In Vitro Selection of RNA Aptamers That Bind Human LysRSΔN65

Honors Research Thesis

Presented in partial fulfillment of the requirements for graduation with honors research distinction in Microbiology in the undergraduate colleges of The Ohio State University

by

Andrew Curtright

The Ohio State University
June 2012

Project Advisor: Professor Karin Musier-Forsyth, Department of Biochemistry and Chemistry

Abstract

Upon infection, HIV-1 begins to reverse transcribe its RNA genome into DNA utilizing $tRNA^{Lys,3}$ as the primer for this process. Though found within the host, it has been shown that HIV-1 requires $tRNA^{Lys,3}$ to be packaged at the time of assembly in order for budding virions to become infectious. Packaging of $tRNA^{Lys,3}$ occurs through an interaction between the lysyl-tRNA synthetase (LysRS)-$tRNA^{Lys,3}$ complex and the HIV Gag-GagPol complex. Inhibition of the host-viral protein interaction occurring between LysRS and Gag presents a potential anti-retroviral strategy. The goal of this project is to discover RNA molecules or "aptamers" that bind LysRS thereby disrupting the interaction with HIV Gag. Systematic Evolution of Ligands by EXponential enrichment (SELEX) is being used to impose an artificial selection on a pool of randomly constructed RNAs. By iteratively separating bound RNA from free RNA, the pool is progressively enriched with molecules which possess a high affinity toward the target protein. After seven rounds of selection, a pool of random RNA demonstrates an enhanced ability to bind LysRSΔN65. Sequence analysis of the pool reveals four homologous regions conserved between unique sequences. These conserved regions fold into stem-loop structures which may play an important role in specific binding to LysRS. Individually isolated aptamers may prove capable of inhibiting the interaction between human LysRS and HIV Gag.

Introduction

HIV-1 is a retrovirus possessing an RNA genome, which is reverse transcribed into DNA and integrated into the host genome upon infection. The complete viral lifecycle is shown in Figure 1. It has been shown that LysRS and $tRNA^{Lys,3}$ are selectively packaged into HIV during assembly [3, 4]. Though LysRS is required for $tRNA^{Lys,3}$ incorporation, the HIV protein Gag is sufficient for LysRS inclusion into new virions [5]. $tRNA^{Lys,3}$ plays a crucial role in the viral lifecycle, acting as the primer for reverse transcription of the RNA genome of HIV. When LysRS and $tRNA^{Lys,3}$ concentrations are increased in HIV virions, the infectivity of the particle is increased. Conversely, when LysRS is knocked down, HIV-1 shows reduced infectivity [7]. In addition to recruiting $tRNA^{Lys,3}$ into new virions, LysRS also plays a pivotal role in the HIV packaging complex involving Gag, Gag-pol, and the RNA genome [11, 13]. Considering the essential interactions that this protein has with many components of HIV-1, LysRS presents a potential target for anti-HIV-1 therapies. The advantage of targeting LysRS is that it is a human protein which is not susceptible to the fast mutation mechanisms of HIV-1 that are responsible for the evolved resistance against many anti-retroviral therapies.

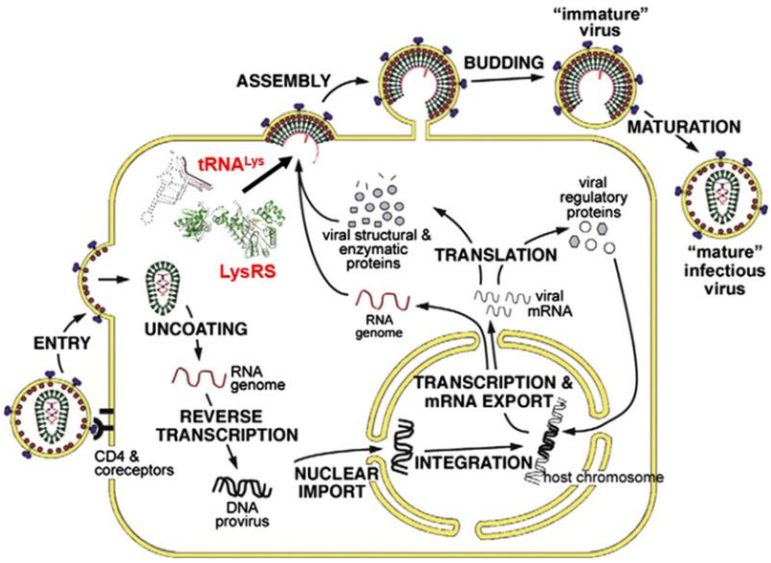


Figure 1: HIV lifecycle [8]

Systematic Evolution of Ligands by EXponential enrichment (SELEX) is a technique developed in the 1990s for isolating small molecules of DNA or RNA against a target protein

[20]. The technique is couched in evolutionary principles: any given population with inheritable variability is capable of adapting in response to a selectional pressure. An extremely large pool of random RNAs can be enriched to possess individual species with binding activity against a target protein through repeated incubation with that protein followed by separation of non-binding molecules from those possessing affinity toward the target. The 'survivors' of the first round of selection are then amplified to regenerate the initial size of the pool before a subsequent incubation with the target protein [20]. After many rounds of selection, the pool can be expected to be composed primarily of RNA sequences with only the highest binding affinity toward the target. Competition between RNA molecules for a limited number of protein molecules is the driving force for selecting only the best binders toward the target.

HIV-1 presents many targets for interference using aptamers derived from SELEX experiments [12]. For example, aptamers have been selected to inhibit the HIV-1 protein reverse-transcriptase [6], as well as HIV-1 Gag's interaction with the RNA genome [16]. Studies have been underway to assess the ability of aptamers to be delivered to infected cells as therapeutics [12]. The SELEX method and associated aptamers present a new hope for novel anti-HIV-1 therapeutics by providing a means for the isolation of small molecule inhibitors without the need for costly design and synthesis of chemical compounds.

In this thesis, we describe our attempt to isolate RNA aptamers to inhibit the LysRS-Gag interaction. An artificially constructed RNA library was evolved to possess binding activity against LysRSΔN65 using two separate methods. LysRSDN65 is a truncated form of LysRS with N-terminal residues 1-65 removed. This region plays a role in non-specific RNA binding and is not a desired target for our selection [18]. The first method utilized native gels to separate the bound from unbound RNA. Later selections switched to using membrane filters of mixed

cellulose esters to achieve separation. Selections using the membrane filters produced an adapted pool with increased affinity to LysRSΔN65. The individual members of this pool were sequenced and analyzed for conserved motifs. Analysis showed that the sequences grouped into four families each with a conserved region. These conserved regions folded into A-rich or U-rich loops and may interact specifically with LysRS.

Methods

Generation of Initial Random Pool

To generate the initial random pool of RNA, a synthetic library of DNA was constructed. The DNA molecules of this library contained a 40 base pair long random region flanked by two primer annealing sites. A 40 nt random region will product 4^{40} unique sequences, all of which cannot be represented in the initial pool since that is equal to 10^{24} sequences or ~680,000 g of DNA. Instead, ~1 ug of DNA of randomized DNA was used to generate an initial pool of ~10^{13} unique sequences. The reverse primer binding region incorporated a T7 promoter site to allow for the transcription using T7 RNA polymerase (*Reverse: CTGTAATACGACTCACTATA*GGGTTCACGGTAGCACGCATAGG; *Forward:* GAGCAGCACAGAGTGGCACAGATGTATGG). The T7 promoter was not transcribed and did not appear in the RNA sequence. This resulted in RNA sequences 92 nucleotides in length, 40 of which were randomized (Figure 2).

5′ – GAG CAG CAC AGA GTG GCA CAG ATG TAT GG – 40N – CCT ATG CGT GCT ACC GTG AAC CC – 3′

Figure 2: Diagram of the DNA template used to generate the round 0 pool

Selection Scheme

DNA was transcribed to RNA using T7 RNA polymerase as described [17]. RNA product from transcriptions was loaded onto a urea gel and visualized using UV shadowing. RNA product bands were excised, crushed, mixed with two volumes elution buffer (0.5 M NH_4OAc; 1 mM EDTA), and incubated overnight at 37 ^oC with shaking. Elution mixtures were centrifuged and the supernatant filtered. Volume reduction was performed via butanol extraction to less than 400 uL. Isolated RNA was ethanol precipitated and re-suspended in nuclease-free water.

Before incubating with protein, RNAs were refolded in 50 mM HEPES buffer and heated to 80 ^oC for 2 min then cooled to 60 ^oC for 2 min. $MgCl_2$ was then added to 10 mM and RNAs were set on ice for 30 min prior to use. Folded RNA was incubated with LysRSΔN65 for 30 min. The bound RNA was then separated from unbound RNA according to two methods described below.

Recovered RNA was amplified using a reverse transcription polymerase chain reaction (RT-PCR) kit from Promega. AMV reverse-transcriptase and Tfl DNA polymerase were used in this reaction. The forward primer included the T7 promoter. The RT-PCR reaction produced DNAs 110 bp in length. DNA was purified with GenElute PCR clean up kit. Purified DNA was used as template for a transcription reaction. The recovered RNA was carried through to another selection. Figure 3 shows a general schematic of the selection.

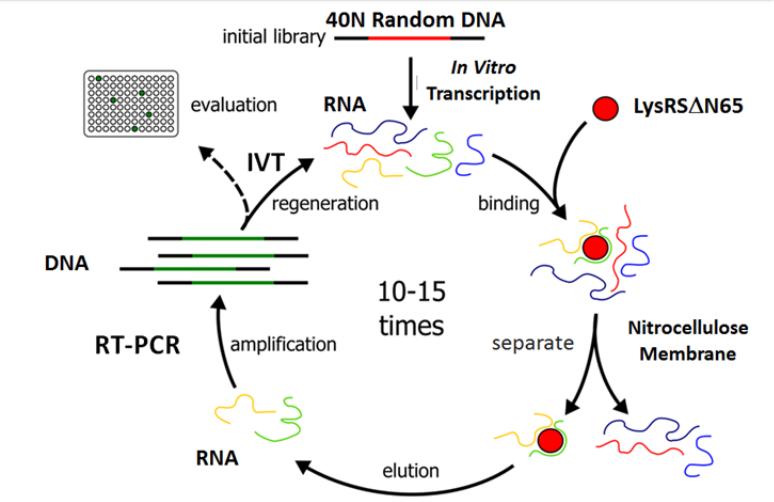


Figure 3: General scheme of RNA selection. Random DNA is first transcribed to RNA and then incubated with the target protein. Bound RNA is separated from free RNA and amplified via RT-PCR to regain the initial size of the pool before beginning another round of selection.

Gel Shift – Method One to Separate Protein-Bound RNA from Free RNA

RNA was incubated with protein in selection buffer (20 mM Tris-HCl; 150 mM NaCl; 1.5 mM $MgCl_2$; 1 mM DTT; pH 7.5) at 37 ^oC for 30 min. The selection pressure was adjusted by varying the amount of RNA and protein as described in Table 1. To increase stringency competitor tRNA was included in later selections. The concentration of NaCl was also varied throughout the selection.

Round	Volume (uL)	RNA (pmole)	Conc. RNA (uM)	LysRSΔN65 (pmole)	Conc. Protein (uM)	RNA:Protein Ratio	NaCl (mM)	tRNA Competitor (ug)
1	45	3,400	75.6	675	15	5	150	
2	45	450	10	45	1.0	10	150	
3	45	450	10	45	1.0	10	150	
4	65	450	7	45	0.69	10	150	
5	65	450	7	45	0.69	10	150	66
6	60	450	7.5	45	0.75	10	250	66
7	60	450	7.5	45	0.75	10	500	66
8	60	450	7.5	4.5	0.075	100	500	66
9	60	450	7.5	4.5	0.075	100	500	66

Table 1: Selection conditions for native gel selections. RNA to protein ratio was increased at round 2 and at round 8. Competitor RNA was introduced at round 5 and salt concentration was increased at round 6.

RNA-LysRSΔN65 incubations were loaded onto a 6% native gel. Samples were run at 4 ^oC at 120V for 2 hours. RNA product was visualized using UV shadowing. Product bands were excised, crushed, mixed with two volumes elution buffer, and incubated overnight at room

temperature with shaking. Elution mixtures were centrifuged, supernatant was filtered and phenol-chloroform extracted to remove residual protein. Recovered RNA was ethanol precipitated and re-suspended in DEPC water. Recovered RNA was used for RT-PCR reactions as described above. Protein-only and RNA-only control lanes were excised and used as template for RT-PCR alongside the selection elutions.

Negative selections were performed by running 450 pmols of RNA (9 uM) on a 6% native gel. Only the lower band was eluted and used in the selection. This removed RNAs with intrinsically low electromobility.

Round	Concentration RNA (nM)	Concentration LysRSΔN65 (nM)	RNA:Protein Ratio	Conc. tRNA Competitor (uM)	$MgCl_2$ (mM)
1	850	850	1		1.5
2	850	850	1		1.5
3	850	850	1		1.5
4	850	283	3		1.5
5	850	283	3	45	1.5
6	1,700	168	10	45	1.5
7	4,200	420	10	45	1.5
8	5,000	50	100	250	1.5
9	5,000	50	100	250	5
10	5,000	50	100	250	7.5
11	5,000	50	100	250	7.5

Table 2: Conditions used in the filter binding selection. The RNA:protein ratio was increased throughout the selection. tRNA competitor was added at round 5 and increased at round 8. The $MgCl_2$ concentration was also increased beginning at round 9.

Filter Binding – Method Two to Separate Protein-Bound RNA from Free RNA

The filter binding protocol used here was adapted from Current Protocols in Biochemistry [15]. Round 0 pool of RNA was constructed by taking 1 ug of randomized ssDNA (about 10^{13} unique sequences) as template for a PCR reaction using 5 cycles of amplification providing approximately 32 copies of each represented DNA sequence.

Selections used SELEX buffer (20 mM HEPES; 15 mM NaCl; 135 mM KCl; 1.5 mM $MgCl_2$; pH 7.2). Before mixing RNA and protein, RNA was refolded as described above, and pre-filtered through membrane filters of mixed cellulose esters pre-soaked in SELEX buffer. Membranes were washed once with 150 uL SELEX buffer. RNAs with ability to bind to the membrane filters without protein were removed from the pool at this point. Flow-through containing RNAs without affinity to the membrane filter were divided into two reactions: one to be incubated with LysRSΔN65 and the other without. Protein or buffer was added to selection RNA and incubated at room temperature for 30 min. Selection reactions were passed through pre-soaked membrane filters and washed three times with SELEX buffer. After the third wash membrane filters were incubated with 200 uL urea buffer (7M urea; 100 mM sodium citrate; 3 mM EDTA) and heated to 80 ^{o}C for 5 min. Elutions were diluted 2x with DEPC water and phenol-chloroform extracted to remove protein. RNA was ethanol precipitated and re-suspended in 50 uL DEPC water. Recovered RNA was used in RT-PCR as described above.

The filter binding method was used through 11 rounds. tRNA competitor was added at round 5 and $MgCl_2$ concentration was adjusted at round 9. Detailed description of the conditions is found in Table 2.

Electromobility Shift Assay

RNA pools were dephosphorylated at the 5' end using calf intestinal phosphatase supplied by New England Biolabs. Dephosphorylated RNAs were incubated with ^{32}P-[γ]-ATP at 37 ^{o}C for 1 hr. Labeled RNA was loaded on urea gel. Radioactive RNA bands were visualized on x-ray film. Product bands were excised and eluted overnight with twice volume of elution buffer. Eluted RNA was butanol extracted and ethanol precipitated.

8000 cpm of labeled RNA (<10 nM) was refolded as described above. LysRSΔN65 was titrated across 50 nM – 10 uM and incubated at room temperature for 30 min. RNA-protein mixtures were run on 8% native gels on ice at 110V for 90 min. Gels were dried and exposed overnight to a phosphorimaging screen and visualized using a Typhoon imager.

Fluorescence Anisotropy

RNA pools were oxidized at the 3' end by incubating 1.5 nmoles of RNA with $NaIO_4$ for 1.5 hours at room temperature. Oxidized RNA was combined with fluorescein-5-thiosemicarbazide overnight at 4 ^{o}C. Excess dye was removed using Roche G-25 Quickspin columns.

Labeled RNA from the round 6 filter binding pool was refolded prior to binding assays as described above. LysRSΔN65 was titrated against 10 nM labeled RNA and incubated at room temperature for one hour. The fluorescence intensity and anisotropy of each sample was measured. Binding assays used SELEX buffer or phosphate buffer (20 mM phosphate, 1.5 mM $MgCl_2$). NaCl concentrations were varied across 0, 50, 100, or 150 mM. pH was also variable from 5.5, 6.0, 6.7, to 7.2.

Radioactive Filter Binding

To optimize buffer pH, filter binding assays were performed using labeled $tRNA^{Lys,3}$. 80,000 cpm of $tRNA^{Lys,3}$ was refolded prior to incubating with 5 uM LysRSΔN65. Mixtures were incubated at room temperature for 30 min. Reactions occurred in phosphate buffer (20 mM phosphate; 15 mM NaCl; 135 mM KCl; 1.5 mM $MgCl_2$ and pH 5.5, 6.0, 6.7, or 7.2). Each mixture was passed through nitrocellulose membrane filters. Membrane filters were washed

three times with buffer. Membranes were dried before taking radioactive counts. The radioactivity of the flow-through was also determined.

Cloning

PCR was used to incorporate primers with flanking restriction sites to the DNA sequences (forward: AATTCC***CTCGAG***GGGTTCACGGTAGCACGCATAGG; reverse: CATTAC***GGATCC***GAGCAGCACAGAGTGGCACAG). *Xho*I and *Bam*HI-HF were used to restrict DNA sequences and the pET vector. Restricted products were purified on 0.6% agarose gel using QiaGen gel extraction kit. Approximately 8 ng of purified insert was incubated with 50 ng of purified vector for a 10:1 insert to plasmid ratio. Ligated product was used to transform competent DH10-B cells via electrophoresis. Transformed cells carried ampicillin resistance as a selectable marker and were plated on LB-ampicillin plates and incubated overnight at 37 ^{o}C.

Isolated colonies were used as template in a PCR reaction to amplify the plasmid area of interest (T7 promoter: CCCGCGAAATTAATACGACTCACTATAGGG; T7 terminator: GCTAGTTATTGCTCAGCGG). These regions were ~750 bp upstream and 50 bp downstream of the insert producing product ~900 bp in length. PCR product was loaded onto an agarose gel to ensure the presence of the proper sized product. Positive reactions were treated with HT ExoSAP-IT® High-Throughput PCR Product Cleanup kit by Affymetrix before being sent for sequencing.

Sequencing and Analysis

Sequencing was performed by SeqWright Inc. Returned sequences were truncated to contain nine bases before and twelve bases after the 40 nucleotide random region. Trimmed sequences were used for alignment. Alignment was performed using ClustalW2. This program

also prepared a phylogenetic tree. Visual analysis was performed to search for conserved sequences present in each phylogenetic grouping.

Prior to subjecting the RNAs to secondary structure folding, the primer sequences were reintroduced to each sequence. Full-length sequences were folded using the mfold web server provided by The RNA Institute at the University of Albany.

Results

Gel Shift – Method One to Separate Protein-Bound RNA from Free RNA

Native gels were used during the selection to separate bound RNA from free RNA. UV shadow visualization was performed to show RNA and protein on gels. A smear of RNA appeared from the top of the gel to the bottom in selection lanes which contained both RNA and protein (Figure 4). RNA aggregates at the top of the gel and near the bottom were denser than the RNA smeared between the two suggesting that these were concentrations of bound and free RNA respectively. Though RNA appeared near the top of the RNA-only control lane, this was not as dense as the aggregates of RNA at the top of the selection lanes. RT-PCR reactions using elutions from controls and selection lanes as template compared relative

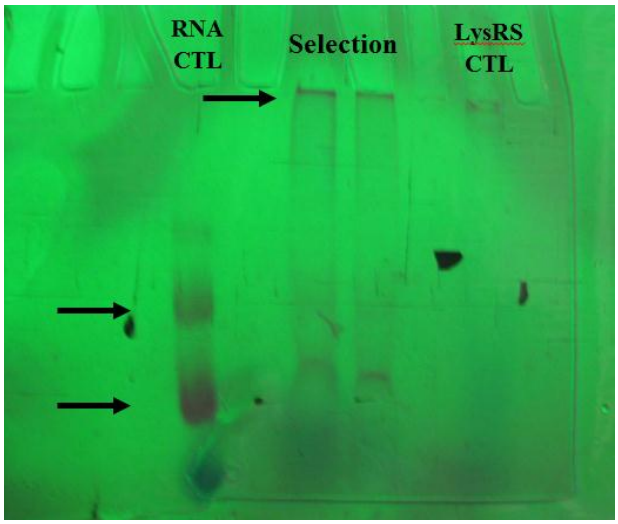


Figure 4: Picture of typical native gel used during selection. Lower arrows point to free RNA migration in RNA control lanes which contained the round X pool of RNA without protein; upper arrow points to RNA-LysRSΔN65 complexes aggregated at top of gel. A faint band of free protein can be seen at the top of the LysRSΔN65-only control lane.

amounts of RNA that was present at the upper region of the gel. RNA-only control lane elutions did not produce as much DNA product as selection lane elutions in RT-PCR amplifications (Figure 5).

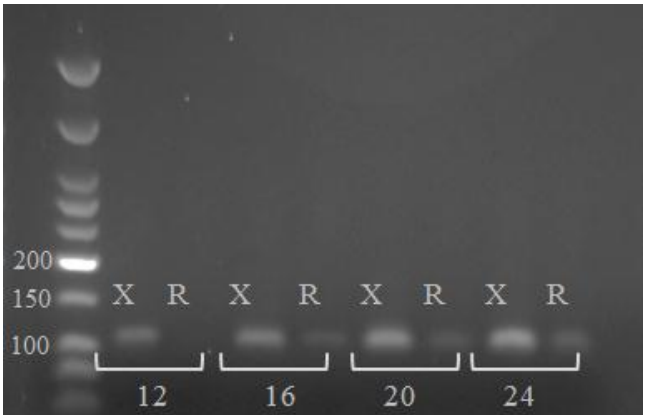


Figure 5: RT-PCR of a typical selection. RNA was eluted from native gel lanes of both the selection (RNA and protein) and control (RNA without protein) reactions and used as template for RT-PCR. Selection lanes (X) show greater RT-PCR product than RNA-only negative control lanes (R) after 12, 16, 20, and 24 cycles of PCR.

The affinity of round 0, 1, 3, and 5 RNA was assayed after five rounds of selection using electromobility shift assays of ^{32}P-RNA labeled RNA pools. The results showed enrichment throughout the five rounds of selection (Figure 6). This enrichment could be seen to a small degree after round 3 and to an even greater degree after round 5.

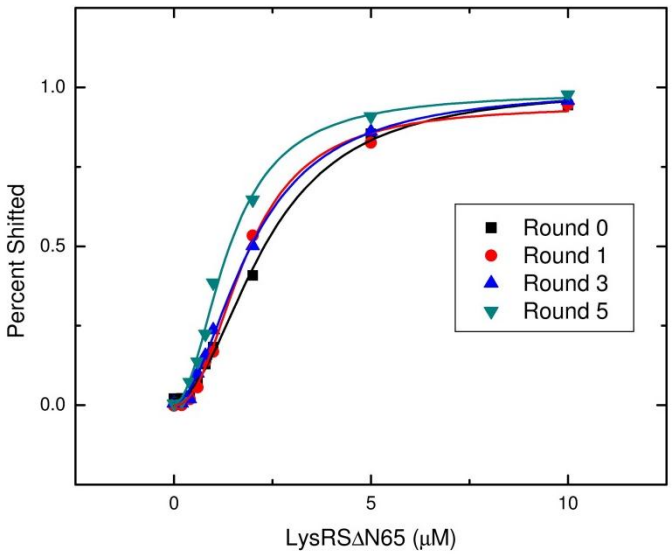


Figure 6: After five rounds of selection electomobility shift assays using ^{32}P-RNA labeled RNA pools from rounds 0, 1, 3 and 5 were performed to measure binding affinity. Each round showed progressively tighter binding to LysRSΔN65.

After round 9, RT-PCR products from selection and RNA-only control reactions suggested that the RNA pool possessed a natural propensity to aggregate at the top of the gel with or without LysRSΔN65 present (Figure 7). A negative selection was performed to remove species which bound to the top of the gel.

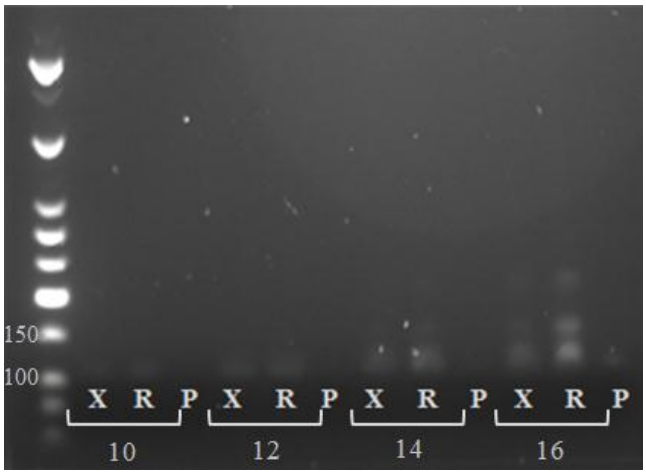


Figure 7: RT-PCR of round 9 selection elutions. "X" contained template derived from gel elutions of selection lanes containing both RNA and protein; "R" from RNA-only lanes; and "P" from protein-only lanes. RT-PCR was performed using 10, 12, 14, and 16 cycles. Gel showed equal amounts of product in both selection and RNA-only lanes.

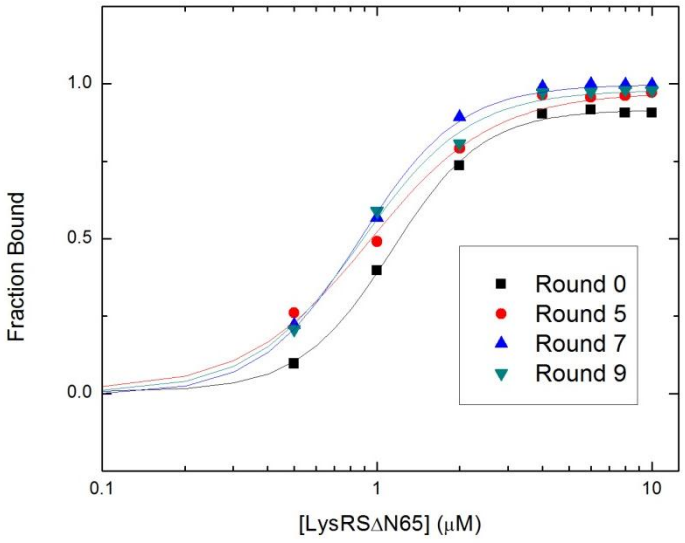


Figure 8: After 9 rounds of selection RNA pools from rounds 0, 5, 7, and 9 were fluorescently labeled. Annealing assays demonstrated an increase in binding affinity toward LysRSΔN65 throughout the selection.

Rounds 0, 5, 7 and 9 were evaluated for their binding affinity toward LysRSΔN65 in EMSAs using radioactively labeled RNA. The assays showed that there was no increase in binding activity from round 7 to round 9 (Figure 9). At this point the selection was terminated and a new method was developed.

Filter Binding – Method Two to Separate Protein-Bound RNA from Free RNA

To determine optimal solution conditions for the filter binding method, ^{32}P-[γ]-ATP labeled tRNA^{Lys,3} was incubated with LysRSΔN65 at various buffer conditions. Reactions were passed through membrane filters and radioactive counts were obtained for each, demonstrating that higher pH will result in less RNA being retained on the filter (Figure 9b). To determine the effect of pH on RNA annealing to LysRSΔN65 tRNA^{Lys,3} was fluorescently labeled and incubated with LysRSΔN65 at different pH points. The affinity of RNA likewise decreased as pH increased (Figure 9a). The decrease in filter retention of radioactively labeled tRNA^{Lys,3} was proportional to the decrease in affinity of fluorescently labeled tRNA^{Lys,3} toward LysRSΔN65 (Figure 9b). These experiments

demonstrated that the decrease in RNA filter retention was due to decreased ability of RNA to bind LysRSΔN65 at higher pH and not reduced filter retention of LysRSΔN65. The effect of NaCl concentration on binding was also determined by performing binding assays at 0, 50, 100, 150 mM NaCl. NaCl was not shown to have a large effect on binding (data not shown).

Fluorescence anisotropy binding assays were performed after the selection had progressed through round 7. Rounds 0, 3, 6, and 7 were evaluated to determine evolution of the pool's affinity toward LysRSΔN65. Round 7 RNA bound better than rounds 0, 3, and 6, with both round 3 and round 6 binding better than round 0 (Figure 10). The increase in affinity suggested that the selection was indeed working. The increase in affinity from round 6 to round 7 indicates that the pool had not reached equilibrium and was still evolving toward LysRSΔN65.

After four additional rounds of selection the binding affinity of rounds 10 and 11 were compared with those of rounds 0 and 7. The assays showed every pool possessing weaker binding than had been previously observed with no differentiation between rounds (data not

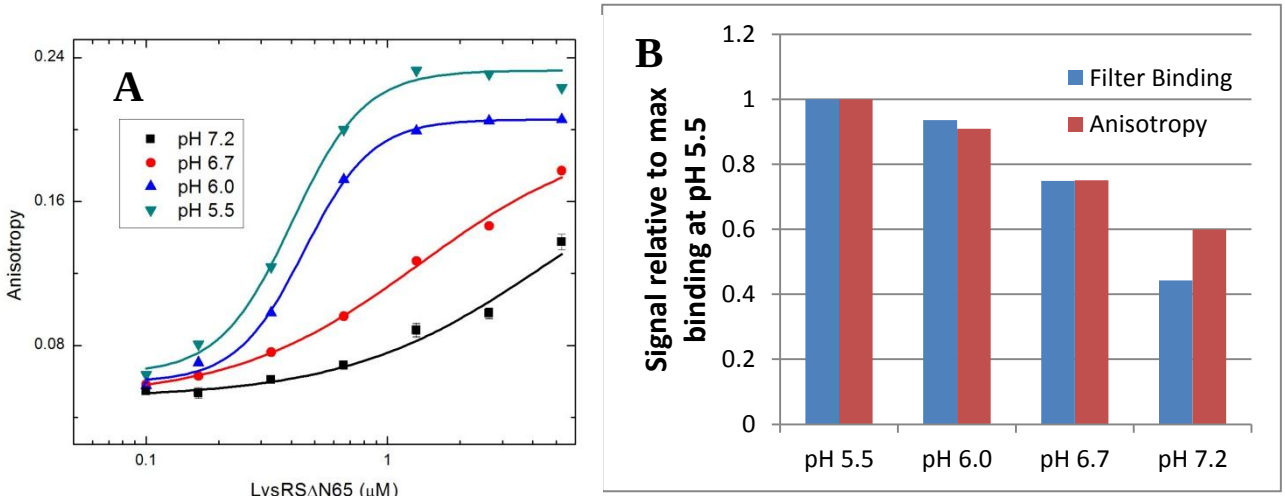


Figure 9: (**A**) $tRNA^{Lys,3}$ was fluorescently labeled and incubated with LysRSΔN65 at different pH points. As pH increased the ability of fluorescently labeled RNA to bind LysRSΔN65 decreased. (**B**) Radioactively labeled RNA was incubated with protein at different pH points and passed through nitrocellulose membrane filters. Radioactive counts of dried membrane filters were obtained and compared to the starting amount of radioactivity in each reaction. As pH increased the amount of RNA retained on nitrocellulose membrane filters decreased. The amount of decrease was proportional to the decrease observed in binding affinity of fluorescently labeled RNA when pH was increased. These results suggest that the reduction in filter retention of RNA observed at higher pH is likely due to reduced affinity of RNA for LysRSΔN65 and not reduced filter retention of LysRSΔN65-RNA complexes.

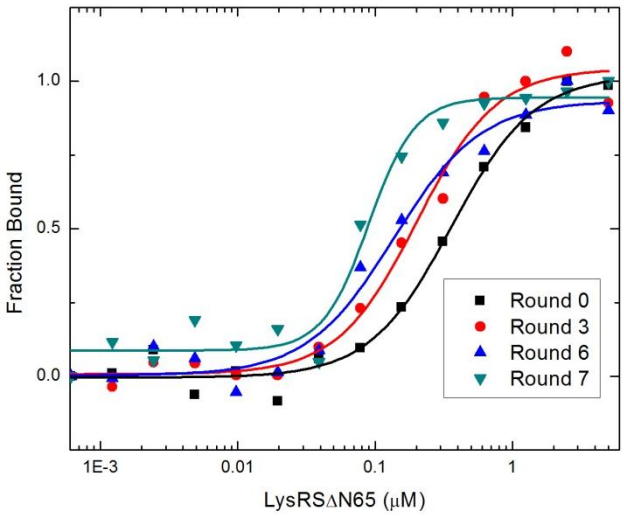


Figure 10: After seven rounds of selection the round 0, 3, 6, and 7 pools were fluorescently labeled and incubated with LysRSΔN65. Annealing assays demonstrate a progressive increase in binding affinity through seven rounds of selection.

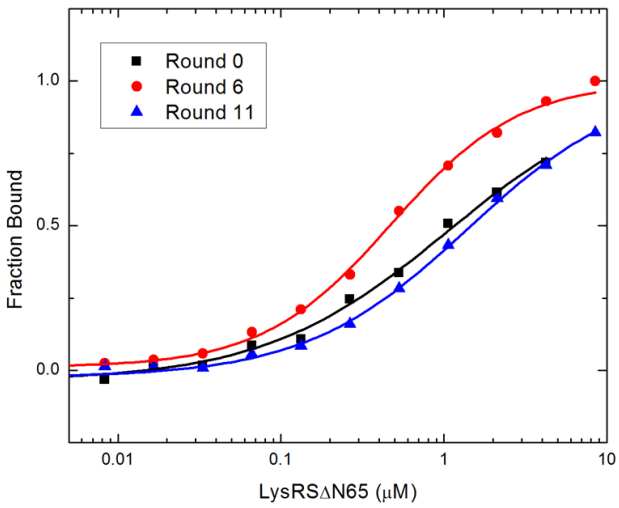


Figure 11: Seven rounds of selection proceeded against properly folded LysRSΔN65 while rounds eight through eleven were against a misfolded form. After 11 rounds of selection rounds 0, 6, and 11 were fluorescently labeled and incubated with LysRSΔN65 in the original conformation. Round 6 demonstrated the best binding against this form of the protein while round 11's binding activity was greatly reduced.

shown). The protein used for the assay was prepared after round 7 and used in rounds 8 to 11. This protein's ability to bind $tRNA^{Lys,3}$ was compared to that of another preparation of the same protein. The assay showed that the protein used for selection from rounds 8 to 11 possessed no binding activity against $tRNA^{Lys,3}$ (data not shown), suggesting that the protein was in an alternate conformation from what had been previously used for the selection. Round 11 RNA was assayed against LysRSΔN65, which did possess $tRNA^{Lys,3}$ binding activity. Round 11 demonstrated less activity than both round 6 and round 0 (Figure 11) indicating a selection away from the target protein.

Sequence Analysis

Of the 96 samples derived from the round 7 pool and sent to SeqWright for sequencing, 72 reads were obtained, 63 of which were not degenerate (Appendix 1). Sequences were aligned

and a phylogenetic tree was constructed (Appendix 2). Visual analysis of determined groupings revealed four sets of conserved elements within phylogenetic groupings.

The largest group (green) had the largest conserved sequence (TGGTTGAAATTGACTA). This region was found in 11 unique sequences, 29 sequences in total. The second group (red) contained 14 unique sequences and 15 total sequences, each with the conserved element (CATCTTTT). The third group (blue) possessed another conserved sequence (CTTCTATGGCTT) and contained two unique sequences with 11 total members. The fourth group (black) contained three total sequences, each unique, possessing a conserved region (AAGTAT).

Structural analysis of representative sequences from each of the phylogenetic families revealed that the most energetically stable structure displayed the conserved region in a stem-loop conformation (Appendix 3).

Discussion

Gel Shift

Stringency The first source of error, which could account for the failure of the gel shift method, may have been the low stringency of the selection. RNA was not present at high enough concentrations above protein to ensure that only the best binders would make it through the selection. This presented the further complication of not having a high enough RNA concentration to saturate the binding of LysRSΔN65. This would produce a low yield which could render the selection vulnerable to stochastic effects such as bottlenecks and founder effect.

The bottleneck effect describes a scenario where an event indiscriminately reduces the size of the population leaving survivors that are not representative of those individuals that are best adapted to their environment [15]. This complication compounds through the founder effect as subsequent populations will only be drawn from the initial stochastically selected pool of individuals. Any well adapted individuals removed during the first bottleneck will never be represented in subsequent rounds of the selection.

The relative stochasticity of the selection can be reduced by increasing the selective pressure on the population. Stochastic events, by definition, eliminate individuals from the population regardless of selective benefit. In order for adaptation to occur, the number of ill-adapted individuals removed from the population through a selective pressure needs to be greater than the number of well-adapted individuals removed through stochastic events.

Having an excess of RNA relative to protein creates a scenario where weak and strong binders must compete for a limited number of binding sites ensuring that only the strongest binders will survive the selection. On the other hand, if there are too many binding sites available both strong and weak binders will survive and ill-adapted RNAs will not be removed. Increasing the number of weak binders removed from the pool buffers the population against stochastic events which will remove weak and strong binders alike. The number of surviving RNAs, however, needs to be high enough that stochastic events do not reduce the size of the pool too low, creating a bottleneck. These considerations are also qualified by the practical limitation of on-hand RNA.

Counter-Selectional Pressures The second problem contributing to the failure of the gel shift method was the abundance of counter-selectional pressures. Various aspects of the selection

may have provided advantages to members of the RNA pool on the basis of properties other than ability to bind LysRSΔN65. These could have been factors such as PCR bias, or ability to aggregate near the top of the gel without protein. These counter-selective forces were evidenced by the large background RNA concentration observed in RNA only negative controls (Figure 7).

It is possible that the ability to aggregate *sans* protein had been selected for throughout all nine rounds of selection though the addition of tRNA competitor drastically raised the total RNA concentration and may have provided a more suitable environment for RNA aggregation (Table 1). Aggregatable RNA did not need to reach a size equivalent to the LysRSΔN65-RNA complex, but rather only large enough to become immobile through the gel. A 6% native gel was being used and lower percentage gels became extremely difficult to work with. Therefore, this problem may be unavoidable given large enough RNA concentrations. Future work should consider the total concentration of RNA and the possible effects on RNA aggregation.

The counter-selectional pressure could have been addressed in the earlier rounds of the selection by having negative selections occur sooner and more frequently. This would have ensured that species possessing an ability to aggregate at the top of the gel were removed from the pool before those species could represent the majority of the population, as was likely the case in this selection. It should be kept in mind that any additional selections run the risk of introducing stochastic forces into the selection through population bottlenecks or adding additional, uncontrollable selectional pressures.

Buffer Discrepancies Another difficulty presented by this method was consistency of buffer conditions. The running buffer of the native gels (Tris/boric acid; 0.5 M EDTA; pH 8.0) was not the same as the SELEX buffer. Loading selection samples onto the gel changed buffer conditions

thereby presenting an opportunity for the samples to re-equilibrate before entering the gel. SELEX buffer could not be used as running buffer due to aberrant gel behavior, especially at high salt concentrations (data not shown). The use of a separate running buffer that remained constant throughout the selection potentially negated adjustments to the SELEX buffer conditions, such as alterations in the salt concentration.

Though the method presented some difficulties, it proved to have advantages. The gel shift method allowed for a small amount of protein to be used. This reduced the amount of RNA required for each round of selection. Limitations in the transcription yields made this a relevant consideration. Additionally, this method provided a means for real time analysis of the selection. RNA separation could be visualized during the selection, giving added assurance that the selection was working.

Filter Binding

Recovery A recurring problem with the early filter binding experiments was low recovery of RNA. The low recovery during early rounds of selection was particularly problematic since it made it difficult to increase the stringency of selection, which would further reduce RNA recovery. The stringency of the selection is most commonly adjusted by modifying the RNA to protein molar ratio. This entails either decreasing the amount of protein or increasing the amount of RNA. Since about 1/3 of RNA product obtained from *in vitro* transcriptions was being used in each selection the RNA concentration could not be increased further to any significant degree. The protein concentration couldn't be lowered too much further either since the protein concentration represents the maximum possible yield of RNA during saturated protein binding and there was already too low of a yield.

Protein Excess Since RNA was a limiting reagent and LysRSΔN65 was in greater supply, it seemed feasible to create a selection condition using LysRSΔN65 in excess of RNA. Though SELEX typically occurs in conditions of RNA in excess of protein there is no essential reason why it must be so. As long as the conditions present an opportunity for RNA with higher affinities toward protein to be separated from RNAs that only weakly bind protein, a selection will occur despite relative concentrations. Using RNA in excess provides the advantage of recovering a large amount of RNA, but our selection was not yielding large recoveries. If the same, small amount of RNA could be recovered using a smaller starting amount of RNA this would conserve reagents and be preferable. Of course, this would reduce the stringency by increasing the amount of survivors *relative to* the starting population, thereby decreasing the selective advantage of RNA binding species. This is not a concern, however, if the selection remains strong enough to still eliminate a large majority of the starting population.

On the other hand, the reduced stringency could actually be beneficial as it would preserve more diversity within the pool. Consider two selections each yielding 1 pmol of RNA; selection A begins with 100 nmols of RNA (100,000-fold more than recovered), selection B with 100 pmols (100 fold more than recovered). If each pool has the same initial diversity (e.g., 1,000 different sequences) then selection A began with 100 pmols of each sequence and selection B began with 100 fmols of each sequence. If 1 pmol of RNA is recovered, then selection B cannot return any fewer than 10 unique sequences, whereas selection A could easily return 1 pmol of the same sequence. Having too many copies of the same sequence may reduce diversity and this problem could be overcome by keeping the copy number low by using a method that uses protein in excess.

Effect of Ligand Concentration on Target Binding To accurately measure dissociation constants (K_D) of RNA-protein complexes, the target should be titrated a log value above and below the K_D, while the ligand must be present at a concentration lower than the lowest concentration of target [10]. As the ligand concentration approaches the K_D the binding curve will shift to the right to compensate for the increased amount of ligand capable of binding (Figure 12). This scenario occurred multiple times throughout the selection when the protein concentration exceeded the K_D of the RNA-LysRSΔN65 interaction and shifted the binding curve right. This resulted in much less binding than expected. At a 5:1 RNA to protein ratio it was expected that the protein would be saturated, however this could not be the case. For saturation to occur the RNA concentration had to be much above the K_D while the protein concentration was much below. During early rounds of the filter binding selection RNA concentrations were not much above the ~300 nM K_D of RNA-LysRSΔN65 observed at the selection conditions, nor were the protein concentrations much below. Counter-intuitively, using less protein may have increased the yield by increasing the fraction of protein bound at a given concentration of RNA.

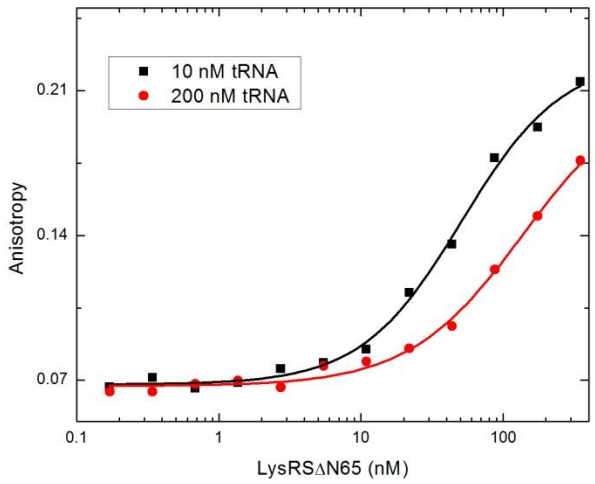


Figure 12: Fluorescently labeled $tRNA^{Lys,3}$ was incubated with LysRSΔN65 at 10 nM and 200 nM. tRNA concentration above the K_D resulted in a rightward shift of the binding curve.

Natural Affinity of the Round 0 Pool The round 0 RNA pool possessing a greater affinity toward LysRSΔN65 than $tRNA^{Lys,3}$ raised questions about the nature of RNA-LysRSΔN65 interactions. The high affinity of the random pool suggests that nonspecific RNA binding interactions are quite strong. It could therefore prove difficult to isolate aptamers which bind specifically, since

specific binding would not have much "selective advantage" over nonspecific binding. It could be compared to the fool's errand of searching for things that stick best to duct tape. Nevertheless, adaptation of the pool could still occur given stringent enough conditions that make it more difficult for nonspecific binders to anneal to LysRSΔN65, but not so stringent to where there are too few RNA species being carried forward to the next round causing a population bottleneck. Nonspecific binding was discouraged by increasing the amount of competitor RNA and using high cationic concentrations to minimize charge-charge interactions.

The homology observed within families of isolated sequences suggests that the selection did not occur on the basis of nonspecific interactions. Had this been the case there should not have been any conserved elements between sequences. Conservation of the same region between otherwise different sequences suggests that the motif provided some sort of selective advantage for those sequences which possessed it. A specific sequence can only provide the advantage of specific binding, for all sequences should provide the advantages associated with nonspecific binding.

Attenuation The decrease in affinity of the round 11 pool toward LysRSΔN65 suggests that the round 7 pool is composed of specific binders to LysRSΔN65. Were the individual RNAs in the pool to bind nonspecifically then selection against misfolded LysRSΔN65 should have had a minimal effect on the ability of the pool to bind to the original form of the protein. The pool would be composed of RNAs able to bind to a protein regardless of structure (e.g. electrostatic interactions) therefore removing the structure should have no effect on binding. This was not the case, as described above, and the pool actually decreased in affinity.

As the pool evolved toward one conformation of the protein it necessarily evolved away from another conformation. This is an example of antagonistic pleiotropy demonstrated in aptamers. An antagonistic pleiotropy occurs when an individual can optimize for one behavior only at the expense of another. This is the principle behind the attenuation of pathogens for vaccines [1].

Sequencing Data

Sequencing data strongly suggested that evolution had taken place throughout the selection. Despite a large variety between sequences, homology was observed. Four conserved elements were maintained in multiple unique sequences suggesting that those regions played an important role in specific binding to LysRSΔN65. Structural analysis revealed that the conserved element of the green family of sequences appeared in an A-rich loop of most full-length RNA folds (Appendix 3). Almost all of the folds created from the red family of RNAs had loops containing UCUUU which contains the $tRNA^{Lys\ 1,2}$ anti-codon CUU (Appendix 4). The blue family also contained loops with the CUU (or UUC) motif (Appendix 5). The black family displayed much less consistency between sequences and folds, displaying no obvious pattern (Appendix 6).

$tRNA^{Lys,3}$ possesses a U-rich loop and is recognized by LysRS (Figure 13). The HIV-1 genome, depending on the specific isolate, possesses either an A-rich or U-rich loop near the primer binding site of the RNA genome which has been hypothesized to interact with LysRS (Figure 13). The selection for U-rich loops of the red family is consistent with the affinity which LysRS demonstrates toward that structural feature in $tRNA^{Lys}$ and the HIV-1 RNA genome. The selection for A-rich loops supports the hypothesis that LysRS possesses a specific affinity for this motif as well. This would suggest that the region of the HIV-1 RNA genome highlighted in

green plays a role in specific binding to LysRS (Figure 13). This is supported by the notable similarity between part of the green family conserved element (UGAAAUU; Appendix 2) and the green highlighted region of the NL4.3 isolate (UGGAAAAU; Figure 13). Further work needs to be done to verify the affinity of sequences possessing these A-rich and U-rich loops toward LysRS.

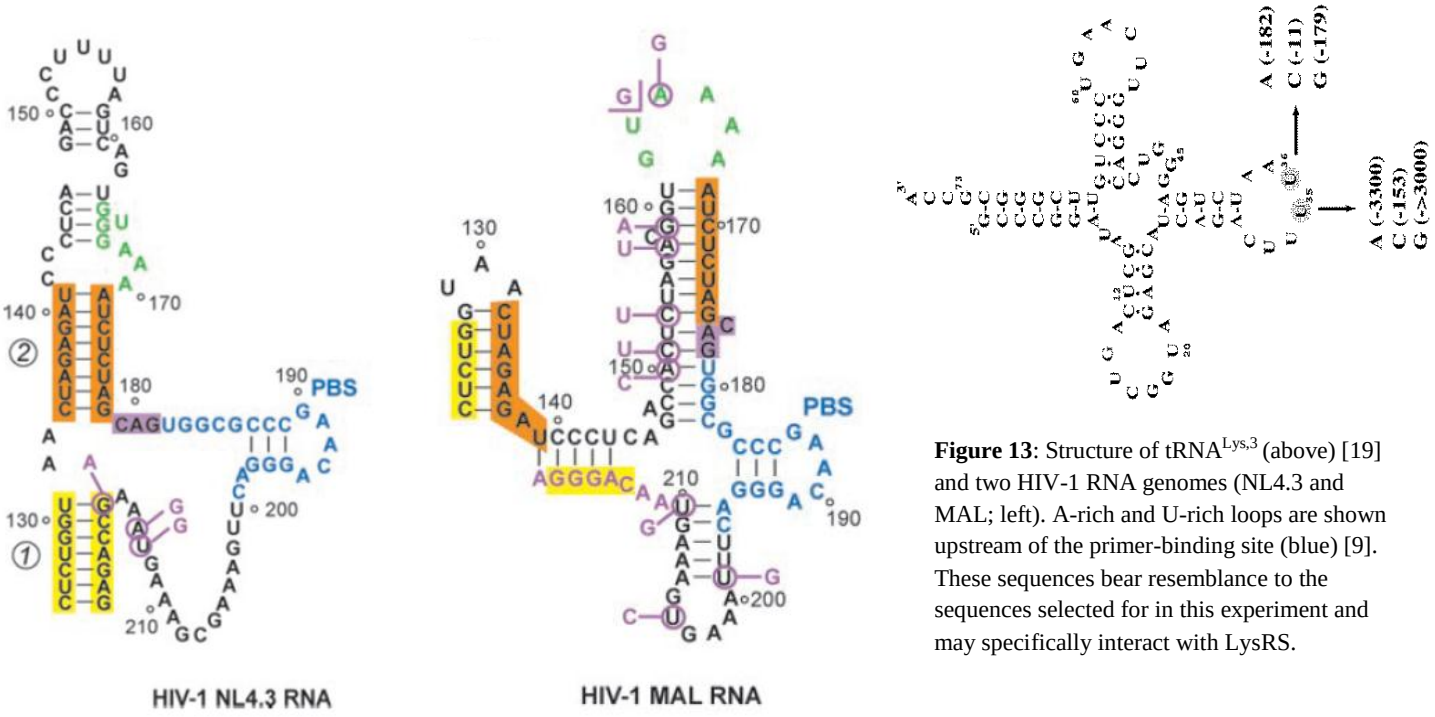


Figure 13: Structure of $tRNA^{Lys,3}$ (above) [19] and two HIV-1 RNA genomes (NL4.3 and MAL; left). A-rich and U-rich loops are shown upstream of the primer-binding site (blue) [9]. These sequences bear resemblance to the sequences selected for in this experiment and may specifically interact with LysRS.

Problems and Difficulties

PCR Artifacts One of the recurring problems with the selection was the appearance of a large PCR artifact. This band showed up during gel analysis of RT-PCR amplification products, as well as during cloning. The band consistently appeared at the same spot whenever the PCR cycle number or template concentration was too high. As the PCR cycle number increased, the proportion of the upper band product to the correctly sized lower band increased. The upper size

band would increase in intensity while the lower sized band would decrease. When PCR product which produces two bands on an agarose gel was used a s template in a transcription reaction, the recovered RNA always produced one band. When each band is excised separately, restricted with *Xho* and *Bam*, and again run on an agarose gel the two bands reappear (Figure 14), even though no PCR reaction was performed in between. Further, the proportion of upper to lower band is about the same regardless of whether the DNA was eluted from the upper or lower band. These previous considerations suggest that the upper band is in fact composed of the same DNA as the lower band. What is causing the shift in behavior remains a mystery though.

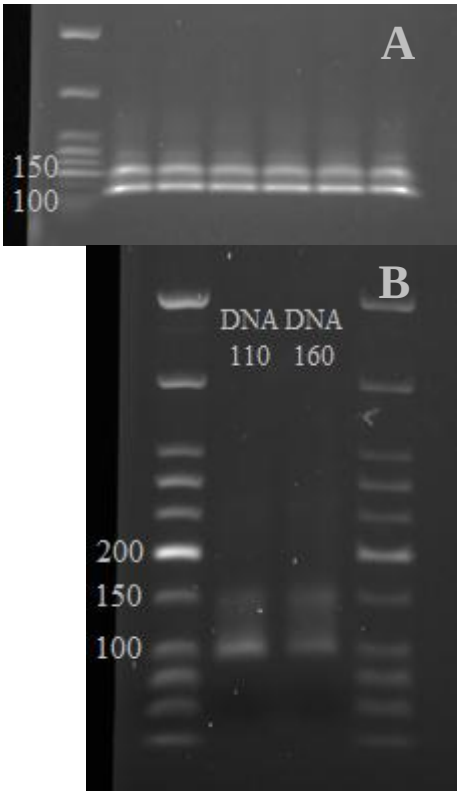


Figure 14: (**A**) Two bands on agarose gel (~110 and ~160 bp) were visible after RT-PCR amplification of round 7 RNA using primers designed to introduce restriction sites. Bands of these two sizes consistently showed up throughout the selection when the PCR cycle number or template concentration were too high. (**B**) DNA eluted from 110 bp length product and 160 bp length product shown in (A) were restricted and loaded onto an agarose gel which showed the same two bands appearing despite template being originally sourced from the upper or lower band.

Background RNA Another problem with the selection was that given enough PCR cycles, we were always able to recover some selection DNA. Selection DNA could be found in protein only negative controls as well as no-template RT-PCR negative controls. Though filtered pipette tips were used during most RT-PCR reactions these problems persisted. This suggests that these small RNA molecules are stable within the lab and care should be taken to prevent contamination of the selection from occurring. This problem was not seen as a large source of error, though it did effectively increase the background noise of the selection. Also these considerations emphasize the importance of having

a parallel negative control to eliminate false positive interpretations where the recovered DNA was sourced solely from the environment and not from the actual selection.

Selection Bias Multiple times the selection was overcome by a large background presence. The above contamination problem was likely a factor but what was likely more serious were the selective forces provided by the method of selection. Regardless of the method, there will always be some selection bias for some RNAs over others. The gel shift method was able to select for RNA that presumably could aggregate together and remain near the top of the gel; the filter binding method could select for RNAs that possessed some affinity toward the nitrocellulose membranes. Given a sufficient number of rounds, these selective forces can substantially impact the selection. Negative selections were performed throughout the filter binding protocol to counter these selective forces, removing RNA molecules which possessed affinity toward the filter before the pool was allowed to interact with LysRSΔN65. Another means by which selection bias could be minimized is by alternating the method of selection every round or two. This would ensure that the only constant selective pressure is that of the protein.

Conclusions and Future Directions

After seven rounds of SELEX a pool of random RNA has been evolved to bind to LysRSDN65. The constituent sequences of this pool group into four families each with a conserved element. These conserved elements fold into A-rich or U-rich loops and may play a role in specific interactions with LysRS.

Much work still remains to be done. Future studies can be performed to determine the binding affinity of individually isolated RNA aptamers. Assaying RNA aptamers which are similar to regions of the HIV-1 RNA genome can provide evidence to evaluate the role of those

regions in binding to LysRS. Specifically, aptamer "Green 59" should be assayed for its ability to bind LysRS and compete $tRNA^{Lys}$. "Green 59" was the most highly conserved sequence which appeared 22 times. Being part of the green family it contains the UGAAAUU region which is similar to the green region of the HIV-1 sequences (Figure 13). This region of the HIV-1 RNA genome has not yet been shown to demonstrate a specific affinity toward LysRS. Since these aptamers were selected as a pool it may be valuable to assay the ability of "Green 59" to compete $tRNA^{Lys}$ in the presence of a red family aptamer which contains the U-rich loop similar to the TLE of the NL4.3 isolate. These two sequences may possess a synergistic ability to compete $tRNA^{Lys}$ greater than the ability for either to compete $tRNA^{Lys}$ alone.

Aptamers possessing the highest binding activity can be characterized further. Assays will need to be performed to investigate the aptamers' effect on the LysRS-Gag interaction as well as these aptamers' ability to interfere with tRNA charging. For those aptamers which inhibit the LysRS-Gag interaction without disrupting tRNA charging, cell based assays can be developed to determine the effect of these aptamers on HIV-1 replication.

The motifs and structures discovered to be conserved among isolated aptamers can be compared to the HIV-1 RNA genome to determine if LysRS interacts with other regions of the RNA genome, specifically regions in the 5' un-translated region. Work in our lab has demonstrated that part of this region, in addition to the TLE, plays a role in specific binding to LysRS (Jones, unpublished data).

The selection could also be taken additional rounds to further enrich the pool. Subsequent selections could include error-prone PCR to introduce variability into the population. This has been performed previously in SELEX [2] and can further enhance the binding ability of

individual aptamers by allowing sequences with high affinities a “second chance” at being selected.

If the selection were to proceed again from round 0 it is suggested that the first rounds of selection occur under much more stringent conditions. This could be accomplished most readily by increasing the RNA:protein ratio. Ultimately, to ensure a strong selection less than 1% of the pool should bind to the target protein. Of course, when RNA returns are so low it is crucial to ensure that the return is significantly larger than the background. Background levels can be estimated using a concurrent negative control which selects for RNAs when no protein is present.

Acknowledgements

I would like to thank Dr. Karin Musier-Forsyth for granting me my first undergraduate research opportunity, and Wei Wang for the support and direction he offered through the good times and the bad. Also, I have the utmost gratitude for all the teachers and professors who have dedicated their lives to empowering students like me to find the truth inside and outside the classroom. Likewise, I am eternally grateful to my loving parents who have always supported me in whatever endeavors I have pursued. Finally, I would like to acknowledge the role the Bahá'í Faith has had in helping me come to appreciate mankind’s miraculous faculty of reason, and its ability to uplift humanity from the depths of chaos.

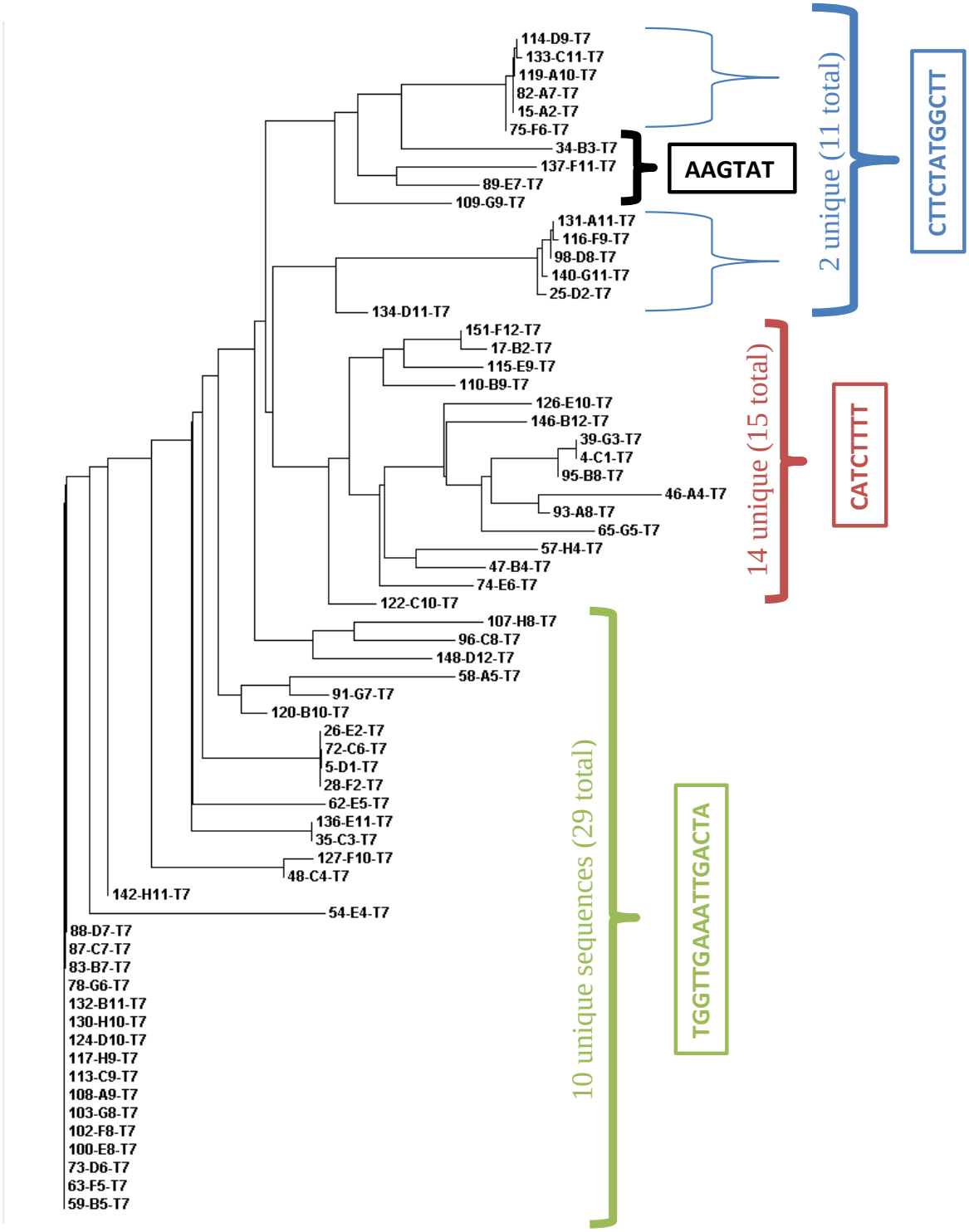
Appendix 1

Clustering analysis performed by ClustalW of all 64 readable sequences.

```
26-E2-T7        TATCGCACATCATGAGCTTGAAA----GACCAATC----TTGT--TGGCG--------- 40
28-F2-T7        TATCGCACATCATGAGCTTGAAA----GACCAATC----TTGT--TGGCG--------- 40
72-C6-T7        TATCGCACATCATGAGCTTGAAA----GANCAATC----TTGT--TGGCG--------- 40
5-D1-T7         TNNCGCACATCATGAGCTTGAAAA---GACCAATC----TTGT--TGGCG--------- 41
110-B9-T7       TACTGT-CG----CGCCTTGGATG---TACGAATCGGAAATCT--TGTTA--------- 40
115-E9-T7       TACTGTATA----GACGTTGAATG----GCCATTCGTATCTTT--TACCG--------- 40
93-A8-T7        --TTA--------GAACTTGAGCC-A-TACGCATC-----TTC--TGCCGCTCTTCGCT 40
58-A5-T7        ------AACGCTGAAA-TTGACTAGG-TGCGTGA--GGTTAGA--TCTTTNC------- 40
91-G7-T7        ----------CTGAAA-TTGACTAGACTGAACGGCTGATGAAA--TATCG--------- 37
20-C2-T7        --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
36-D3-T7        --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
44-H3-T7        --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
38-F3-T7        --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
52-D4-T7        --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
59-B5-T7        --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
63-F5-T7        --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
73-D6-T7        --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
78-G6-T7        --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
83-B7-T7        --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
87-C7-T7        --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
88-D7-T7        --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
100-E8-T7       --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
102-F8-T7       --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
103-G8-T7       --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
130-H10-T7      --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
132-B11-T7      --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
124-D10-T7      --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
117-H9-T7       --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
113-C9-T7       --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
108-A9-T7       --TCCTCTGGTTGAAA-TTGACTA-G-CCCAAAACTTGGAC----TGTA---------- 40
62-E5-T7        -----TCT--TTGAAA-TTGATTATA-CGCTAAACTAACGTA---TGTAGGG------- 40
35-C3-T7        --TTATGGAGTTGAAA-TTGACTA-----CAAAGAT-GCTTG---TGTACG-------- 39
136-E11-T7      --TTATGGAGTTGAAA-TTGACTA-----CAAAGAT-GCTTG---TGTACG-------- 39
48-C4-T7        --ACTTGTGGTTGAAA-TTGACTTGN-AGTACAAC--NNTNC---NGT----------- 39
127-F10-T7      --ACTCGTGGTTGAAA-TTGACTTGG-AGTACAAC--GTTCC---AGTA---------- 40
54-E4-T7        -------CGACCGCAG-GT-AGTGCA-AGTACATCCACTTTAA--GGTGTGA------- 40
107-H8-T7       --ATCNNCGGTCGNTCTNTGAGTT-G-GTTGAAATTGACTA----GATG---------- 41
15-A2-T7        ---TATGTAGGCTC-C---AAGTA---CTATCACTTCAGGTTG--TGACGCG------- 40
82-A7-T7        ---TATGTAGGCTC-C---AAGTA---CTATCACTTCAGGTTG--TGACGCG------- 40
114-D9-T7       ---TATGTAGGCTC-C---AAGTA---CTATCACTTCAGGTTG--TGACGCG------- 40
119-A10-T7      ---TATGTAGGCTC-C---AAGTA---CTATCACTTCAGGTTG--TGACGCG------- 40
75-F6-T7        ---TATGTAGGCTC-C---AAGTA---CTATCACTTCAGGTTG--TGACNCG------- 40
133-C11-T7      ---TATGTAGGCTC-C---AAGTA---CTATCACTTCAGGNTG--TGACGCG------- 40
137-F11-T7      ----TTGTCNNTCCAG---AAGTA---TACGCACTTCAAC-GN--NGAATGGA------ 40
109-G9-T7       ---TTTATTGCTGTAC---AAGTA---TCAGAACTTAATCTGG--TGACTA-------- 40
89-E7-T7        --TGACTTTGGGCC-A---AAGTA---TCCACTCTTTAT-TGG--TGACGGA------- 40
98-D8-T7        ---CTTAT-GTTCCAA---ATGTC---TCCATATCTTCTATGG--CTTTGTT------- 40
131-A11-T7      ---CTTAT-GTTCCAA---ATGTC---TCCATATCTTCTATGG--CTTTGTT------- 40
116-F9-T7       ---CTTAT-GTTCCNA---ATGTC---TCCATATCTTCTATGG--CTTTGTT------- 40
140-G11-T7      ---CTTAT-GTTCCAA---ATGTC---TCCATATCTTCTATGG--CTTTGCT------- 40
25-D2-T7        ---CTTAT-GTTCCAA---ATGTC---TCCATATCTTCTATGG--CTTTGTA------- 40
39-G3-T7        ATTCAAATA----CTGACCAT---GCATCCGCATCTTTTGCC---CATGG--------- 40
4-C1-T7         ATTCAAATA----CTGACCAT---GCATCCGCATCTTTTGCC---CATGG--------- 40
95-B8-T7        ATTCNNATA----CTGACCAT---GCATCCGCATCTTTTGCC---CATGG--------- 40
65-G5-T7        -TTCCGGTA----CTTGC-AT----CATGCGAATCTTCTTCCATGCATTG--------- 40
17-B2-T7        ---TTGATAG---CTCGTGACTG-GTATGCGAATCTTTTT-C---CCCAGC-------- 40
151-F12-T7      ---TTGATAG---NTCGTGACTG-GNANNCGAATCNTTTTTC---CCCAGC-------- 41
126-E10-T7      ---ATAATCA--TTGGGTTCCG---CATACGCATCTTTTGCC---CGTTTC-------- 40
146-B12-T7      -----GGTCA--TATAGGTCCGTTCCATCCGCATCTTTTGCC---NAACA--------- 40
34-B3-T7        -------TAG--TGAGGCTGCGA--CAACGACATCTTTTGCAA--CGGAACAA------ 40
46-A4-T7        -CCCGAGTCACATATAGTAGCT---CATACGCATCTTCTGCC---AG------------ 40
57-H4-T7        --TAACATAA---ATAGGTGGAAA-CGGTCATATCTTTTGTG---AGTT---------- 40
74-E6-T7        ----TTATA----CGGAAAATCG---ATCGACATCTTTTGCTTAAGACGAC-------- 40
```

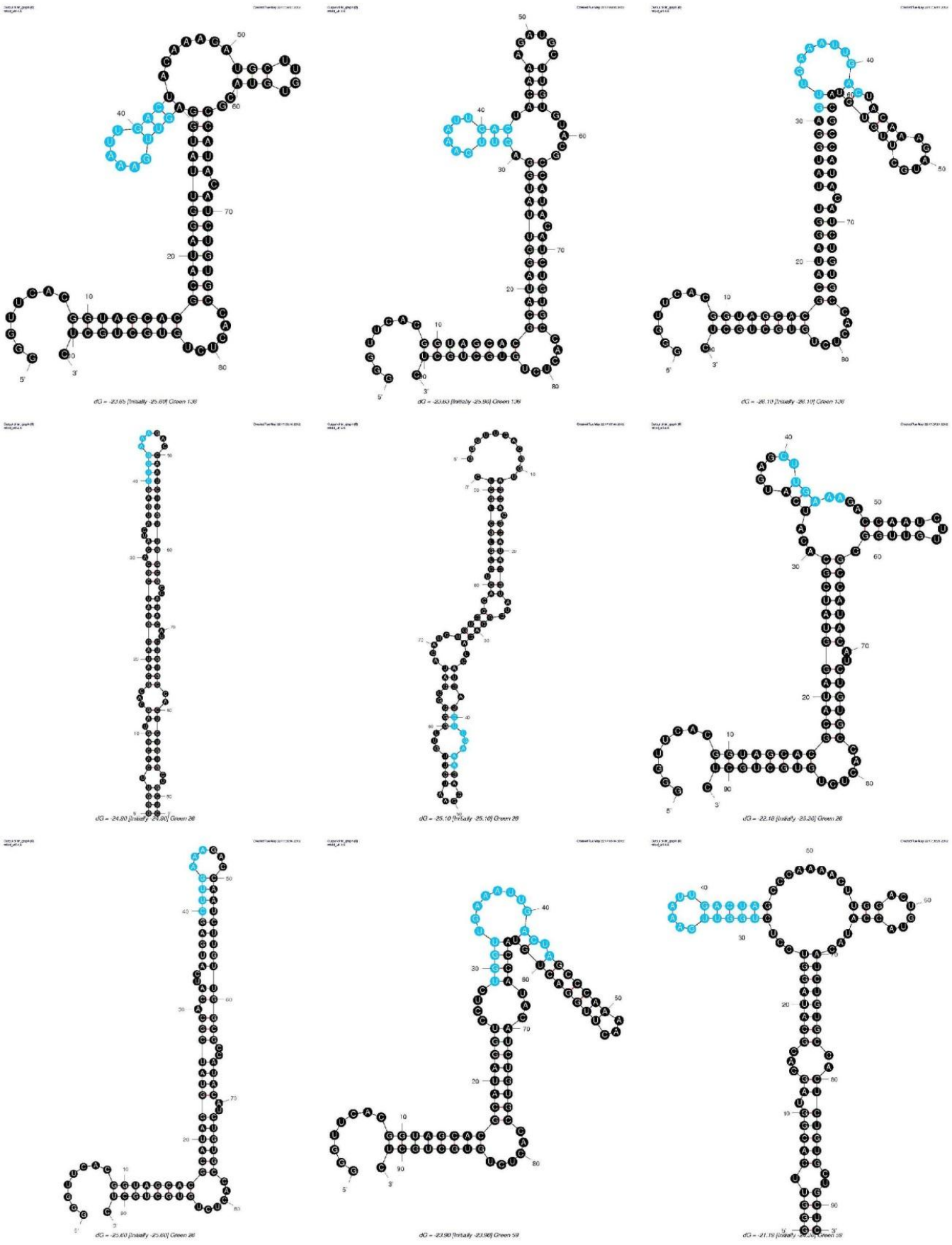
Appendix 2

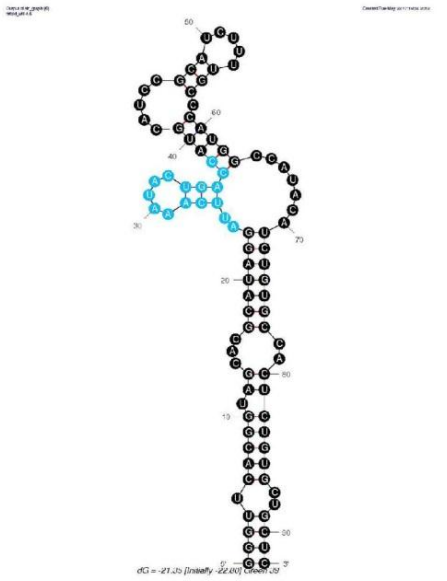
Phylogenetic tree constructed using ClustalW grouped sequences into four families. Each family possessed the conserved region noted. Also listed are the number of total and unique sequences belonging to each family.

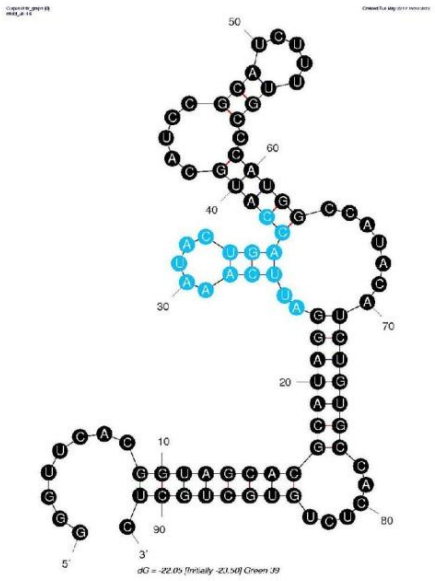


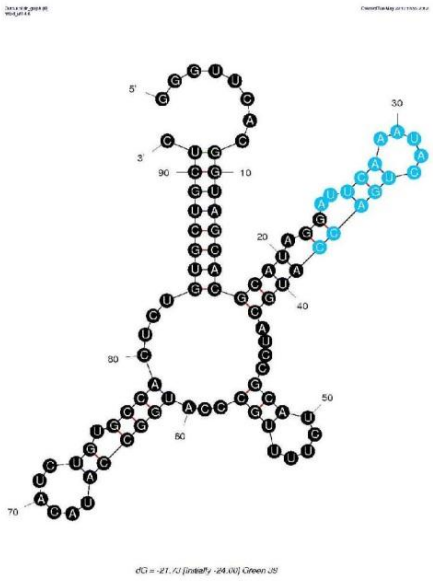
Appendix 3

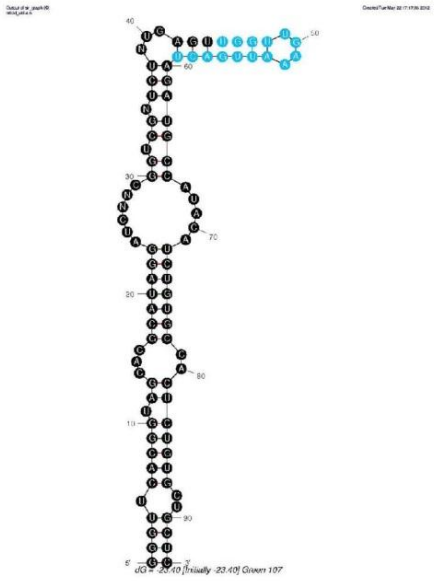
Sequences from the green phylogenetic family were folded using mfold. Conformations with the lowest free energy are displayed. The conserved region is highlighted.

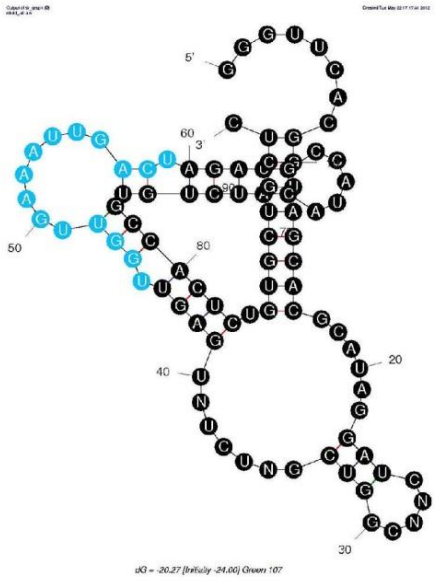


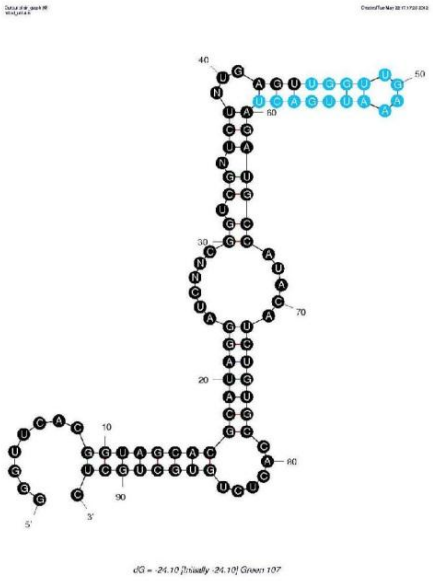


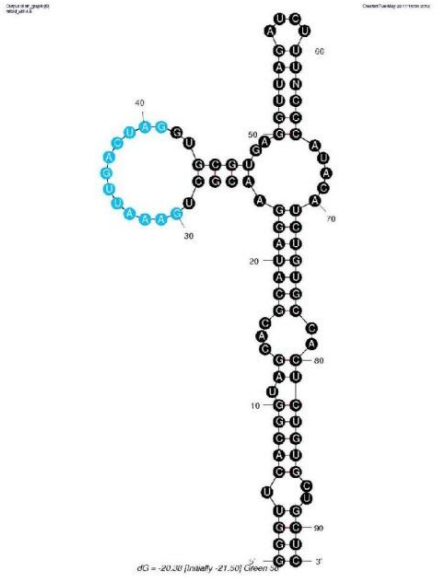


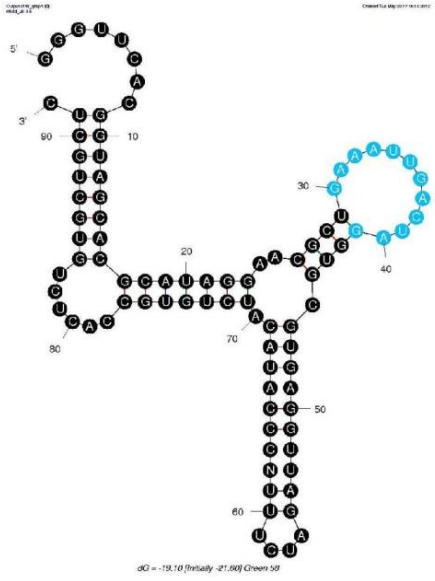


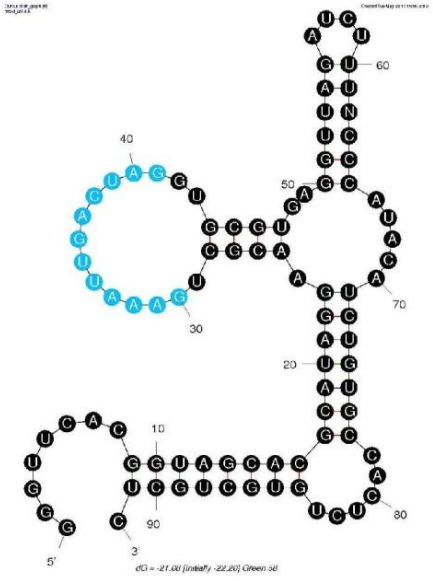


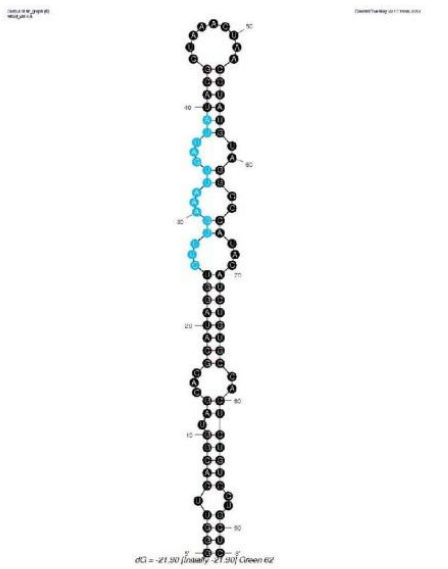




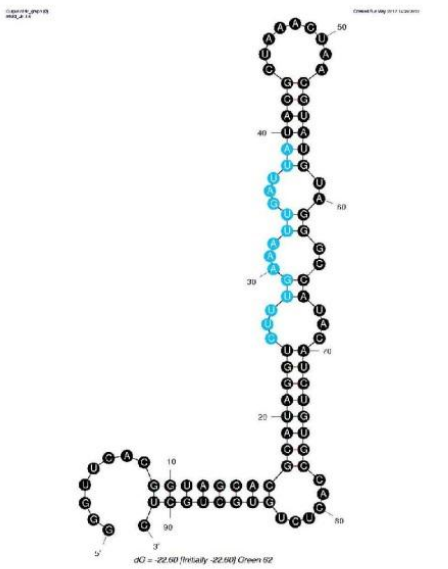




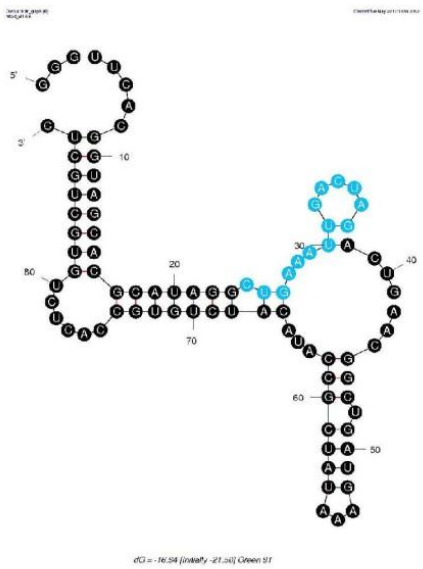




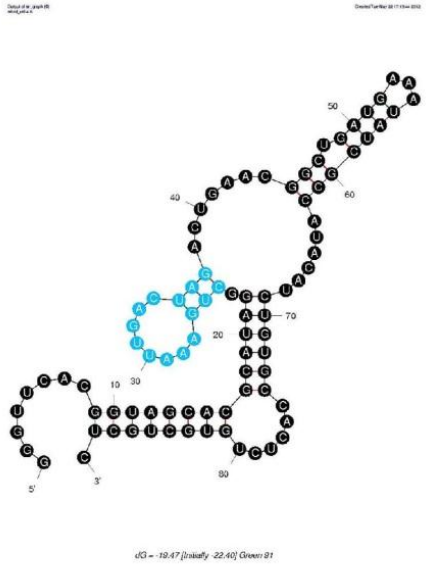
dG = -21.30 [Initially -21.30] Green 62



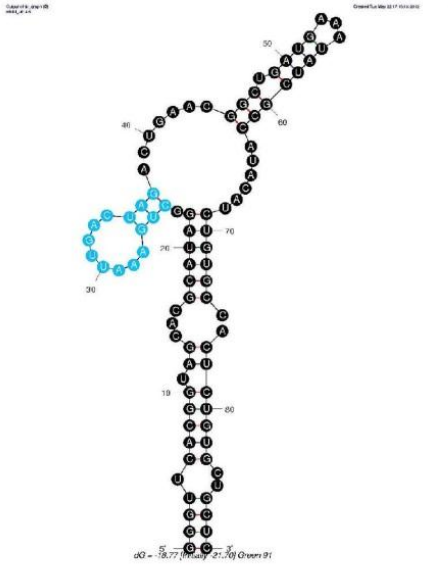
dG = -22.60 [Initially -22.60] Green 92



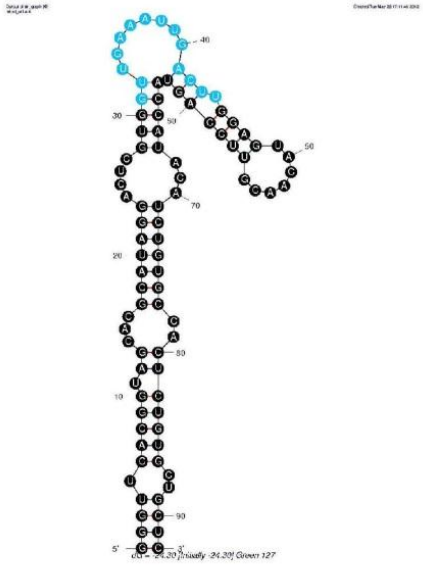
dG = -16.54 [Initially -21.34] Green 91



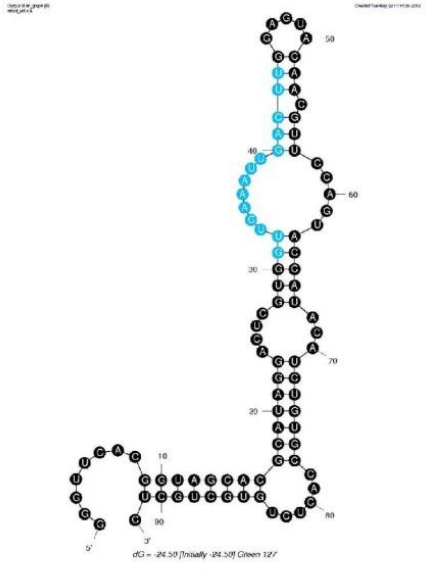
dG = -19.47 [Initially -22.40] Green 91



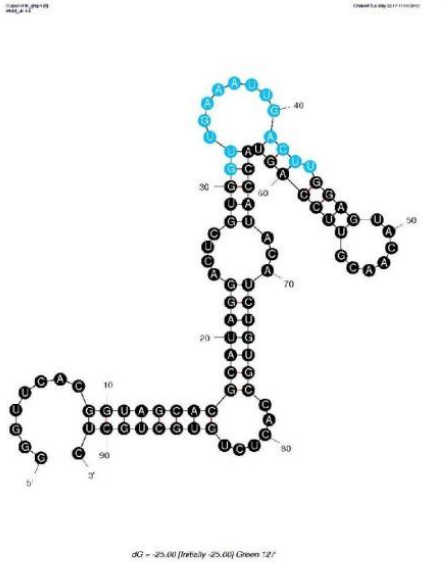
dG = -18.77 [Initially -21.70] Green 91



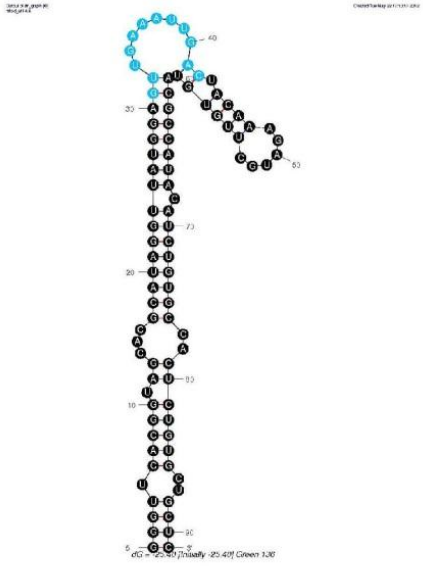
dG = -24.30 [Initially -24.30] Green 127



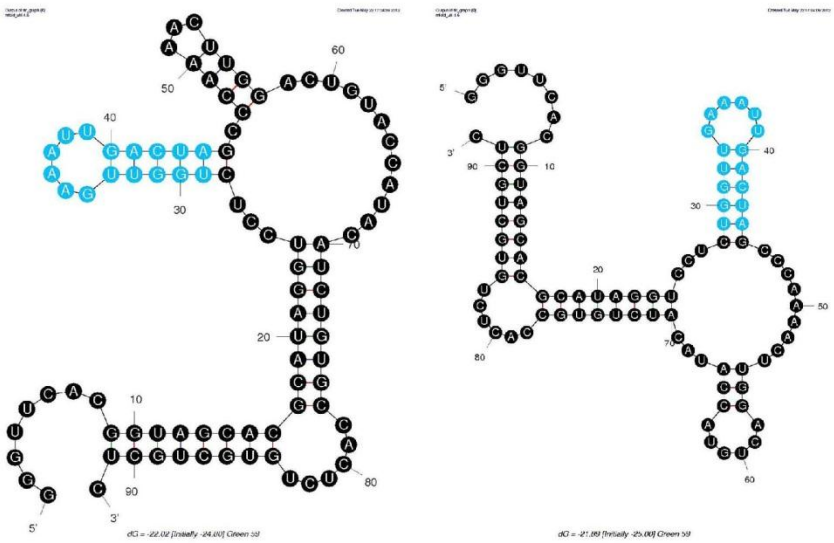
dG = -24.50 [Initially -24.50] Green 127



dG = -25.00 [Initially -25.00] Green 127

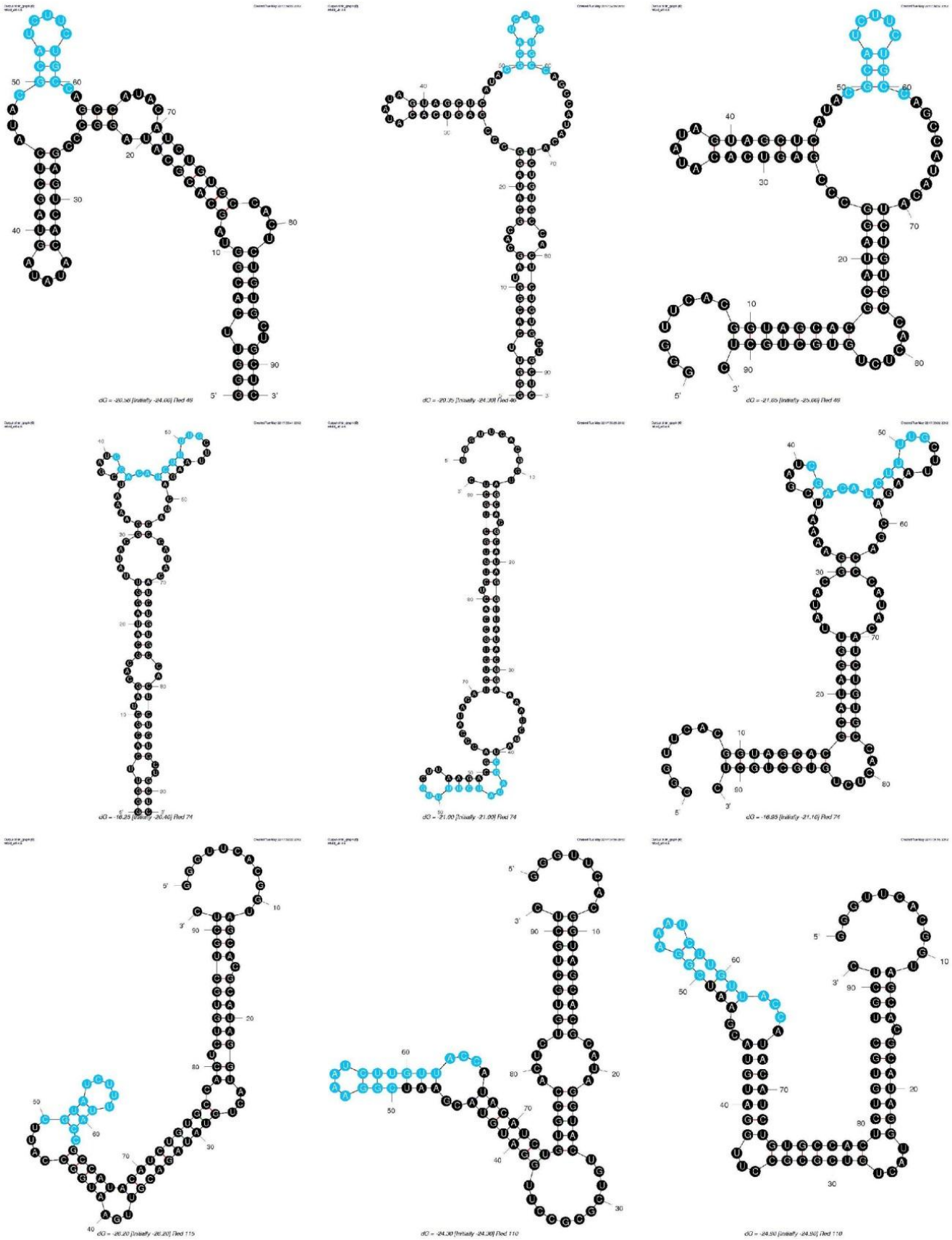


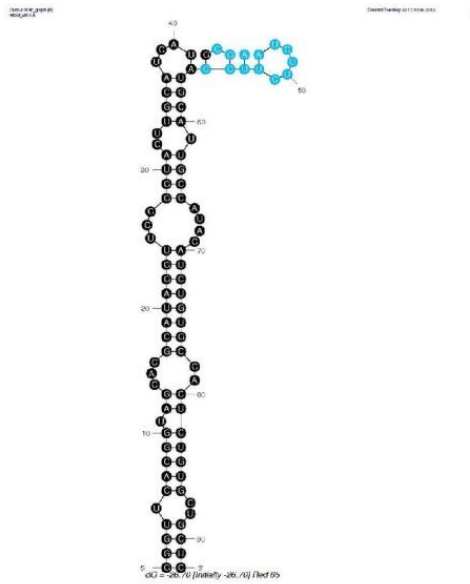
dG = -25.40 [Initially -25.40] Green 136

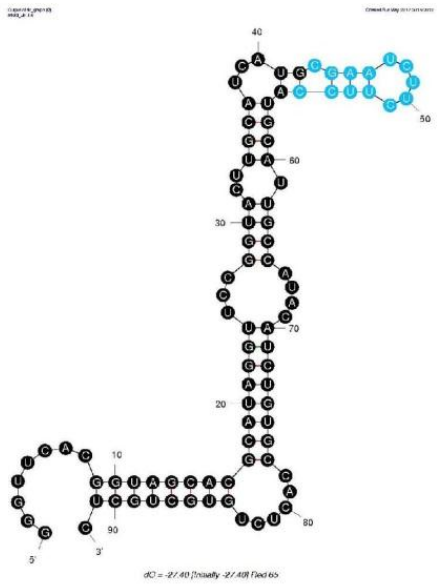

dG = -22.02 [Initially -24.00] (Green 59
dG = -21.89 [Initially -25.00] (Green 59

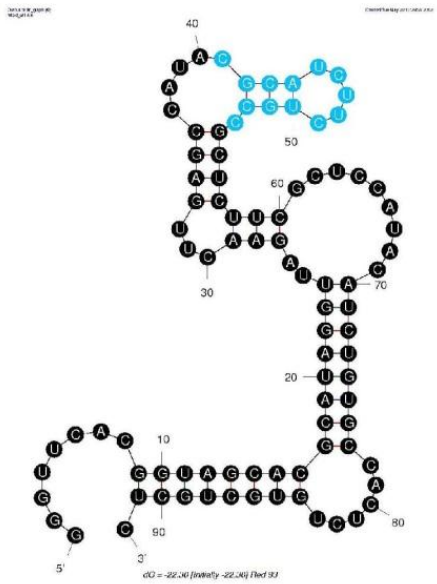
Appendix 4

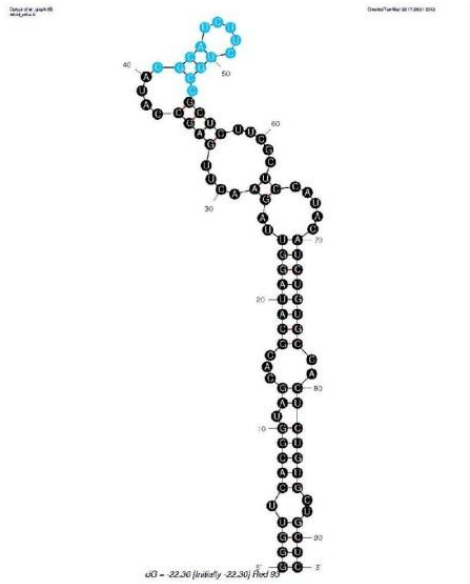
Sequences from the red phylogenetic family were folded using mfold. Conformations with the lowest free energy are displayed. The conserved region is highlighted.

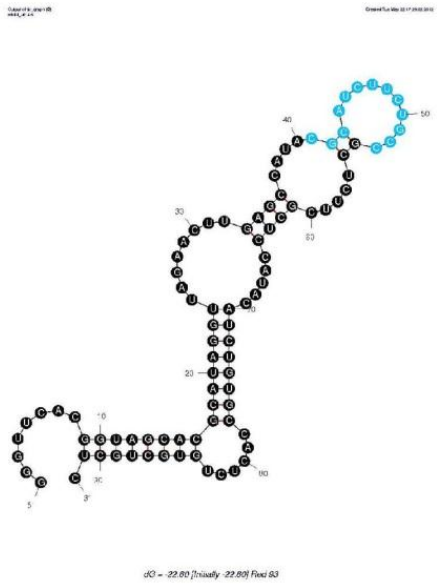


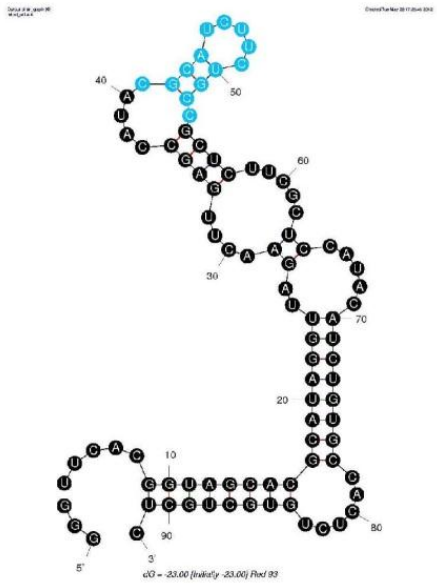


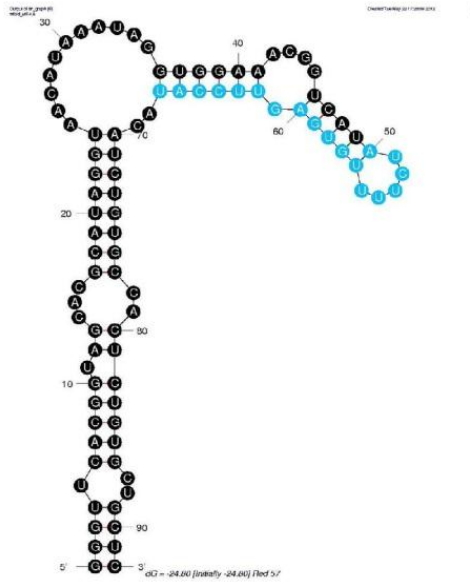


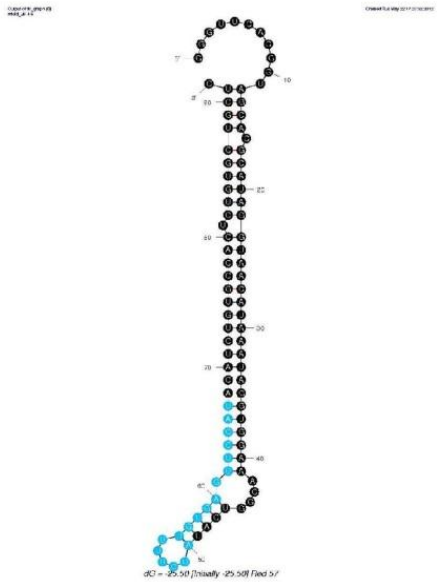


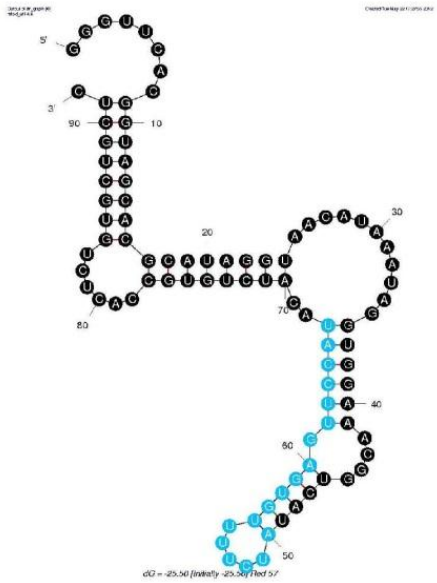


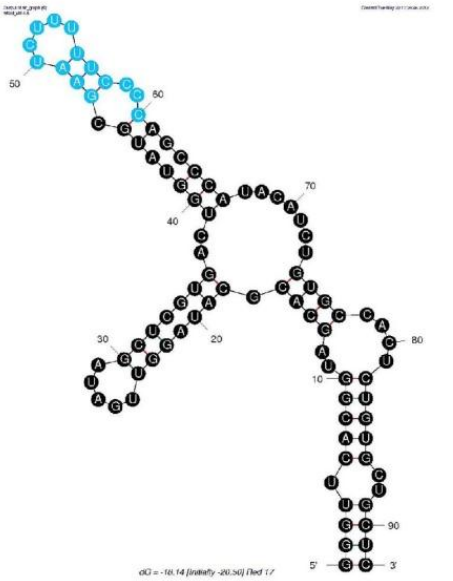


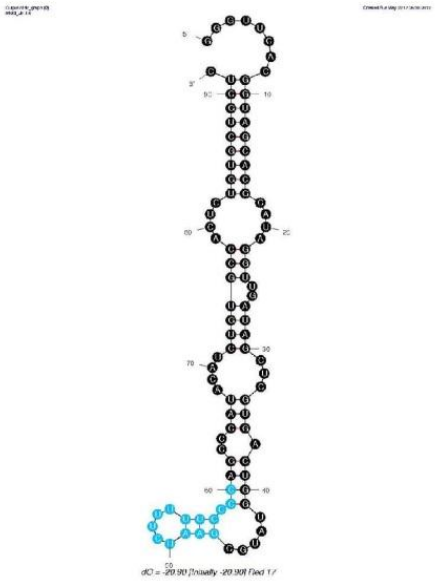


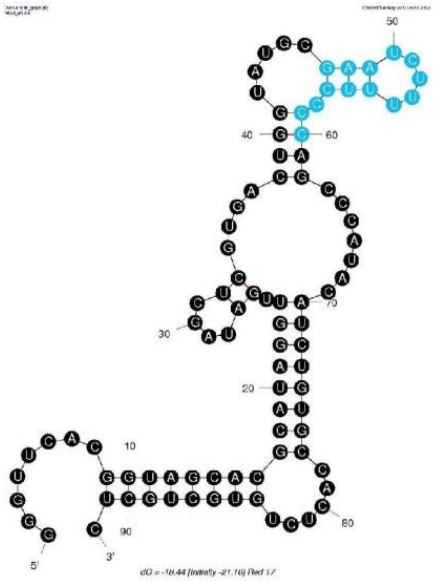


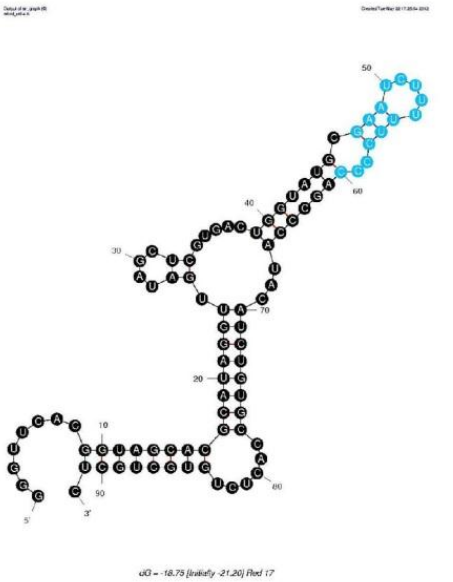


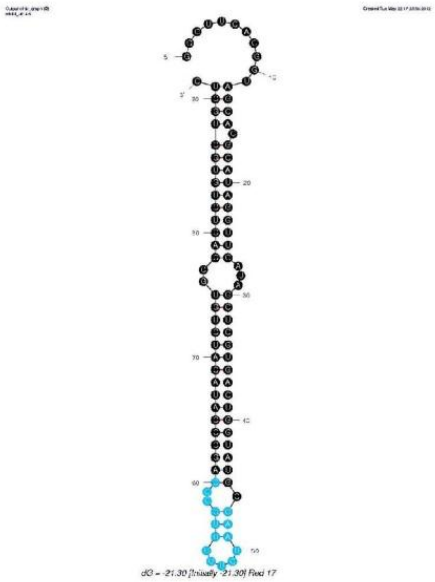


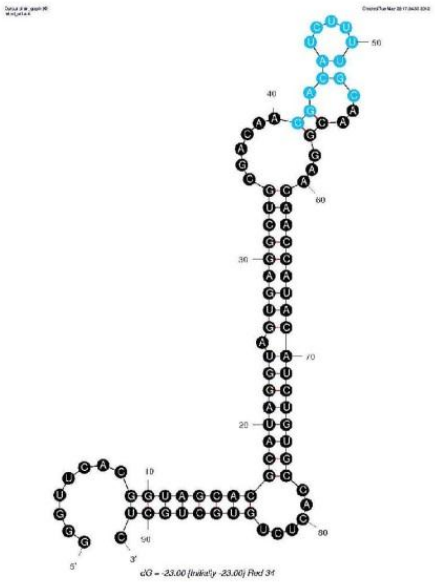


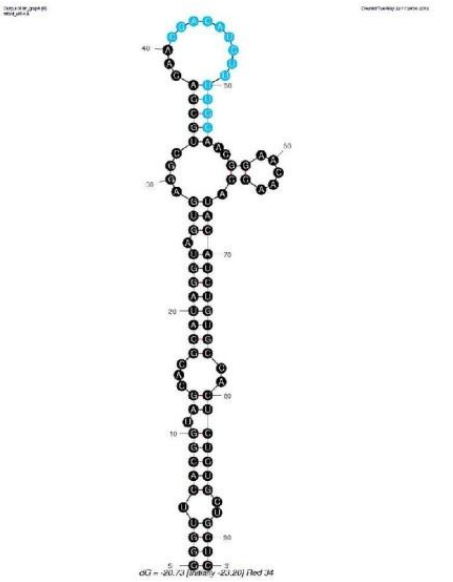


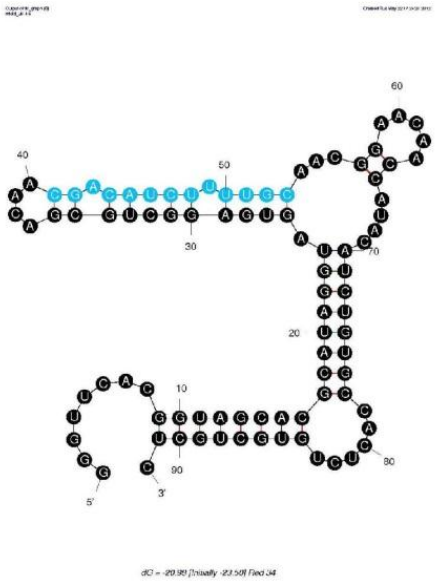


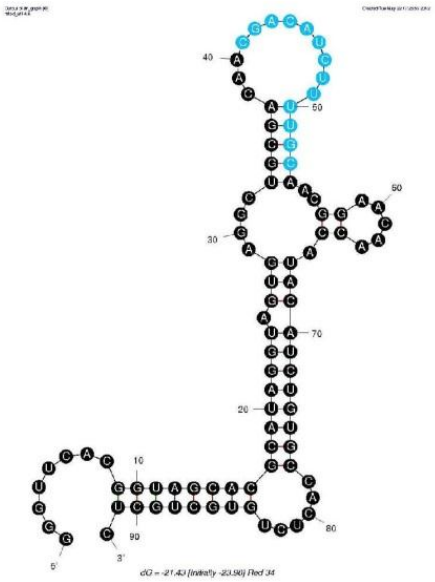


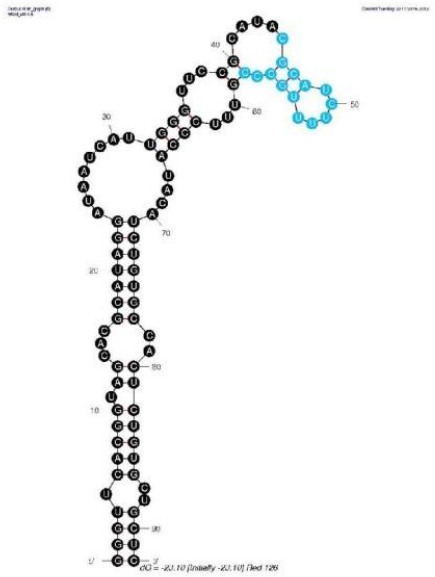


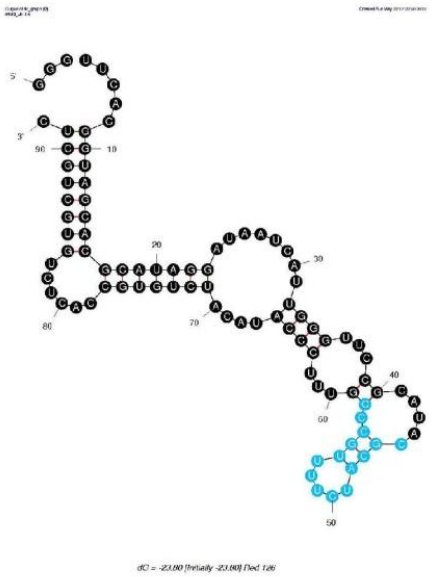


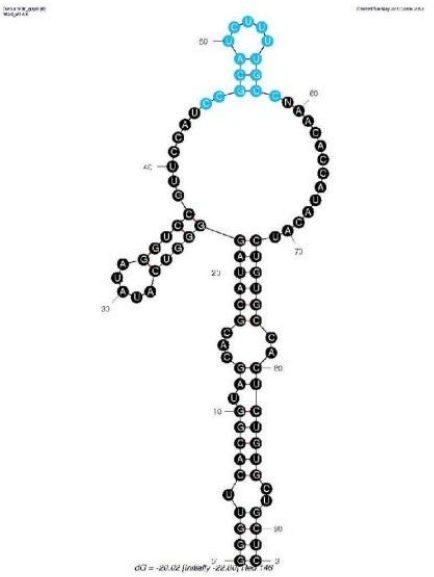


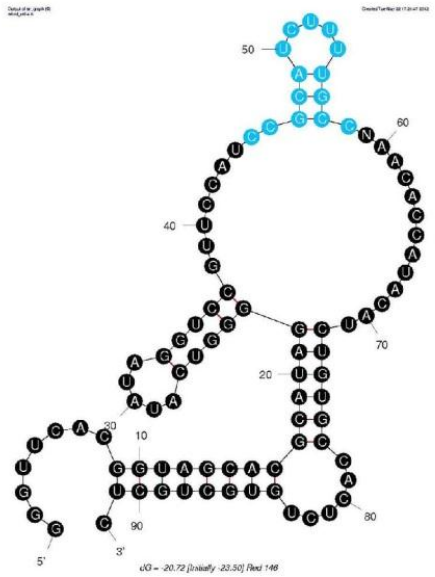


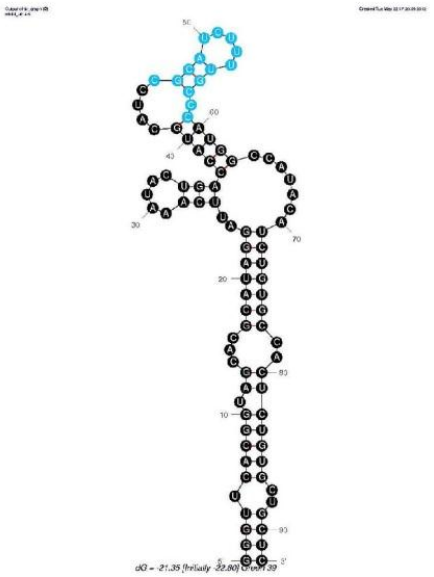


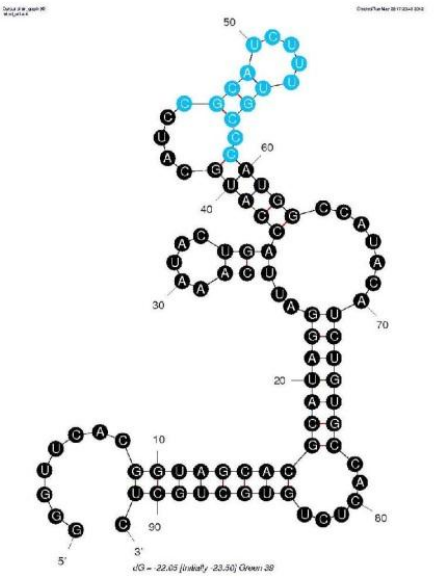


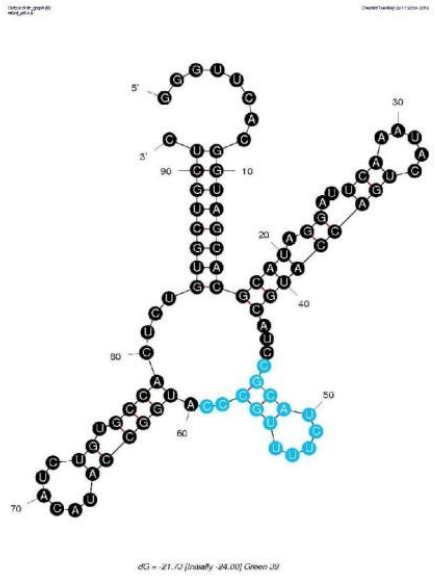






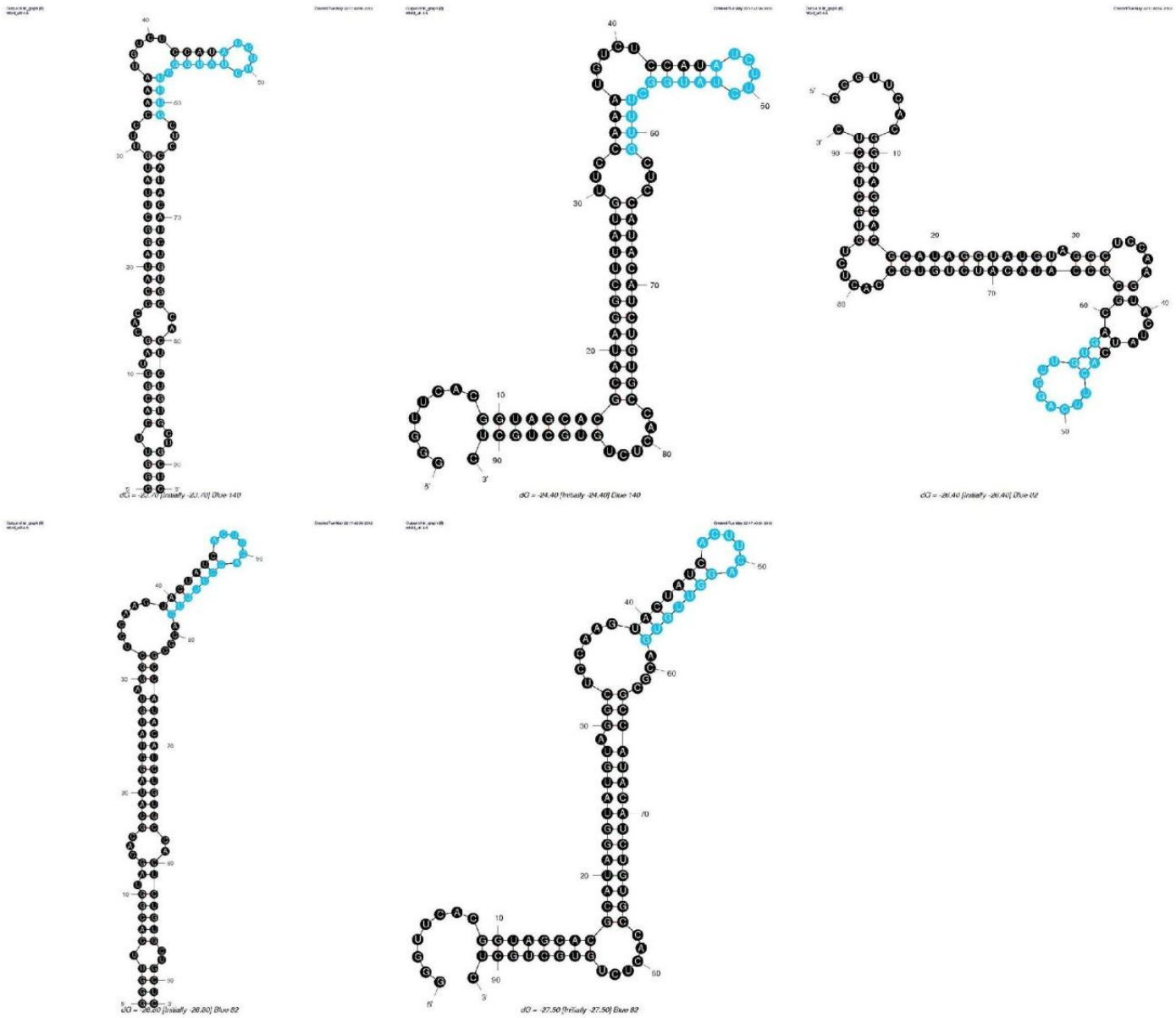






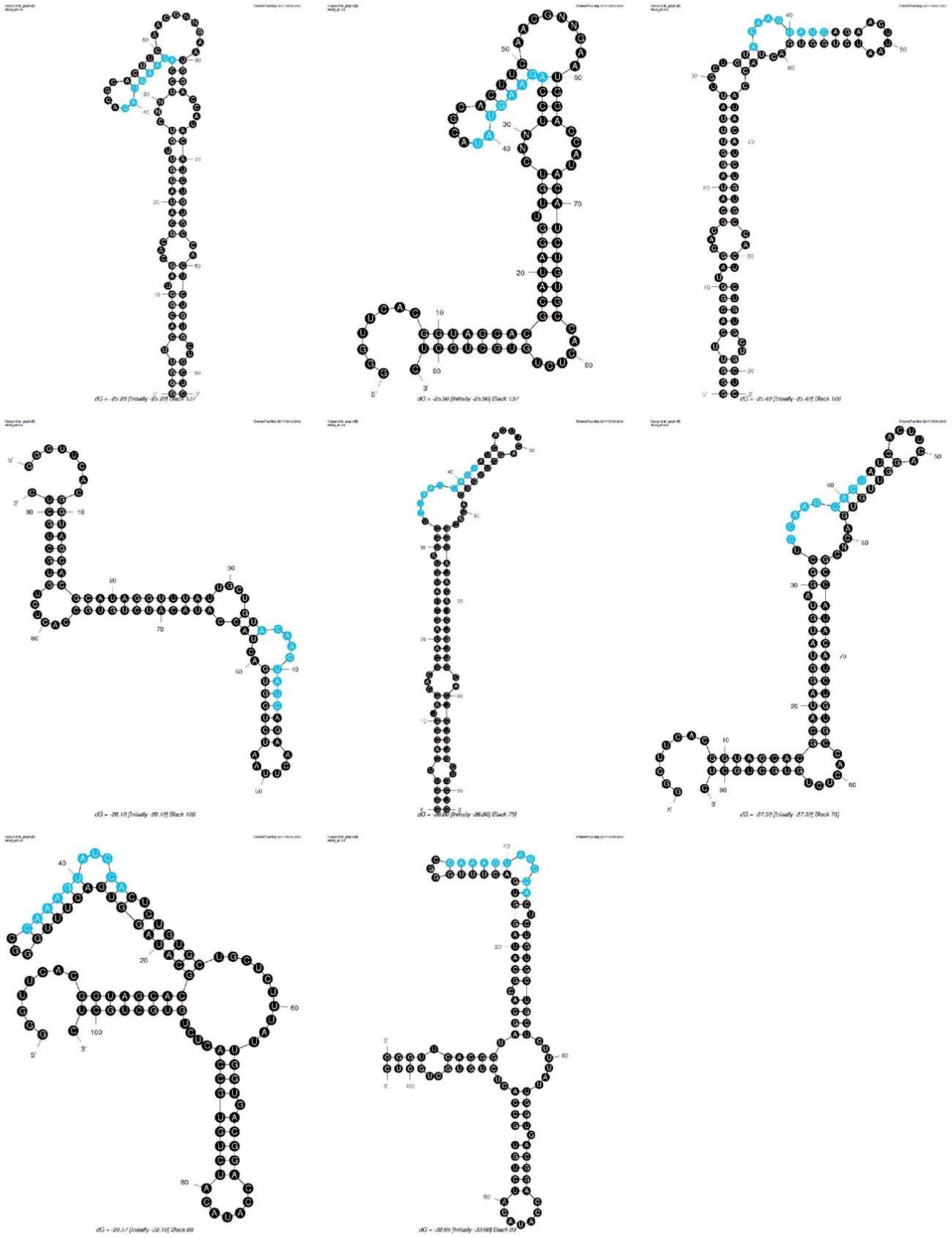
Appendix 5

Sequences from the blue phylogenetic family were folded using mfold. Conformations with the lowest free energy are displayed. The conserved region is highlighted.



Appendix 6

Sequences from the black phylogenetic family were folded using mfold. Conformations with the lowest free energy are displayed. The conserved region is highlighted.



Works Cited

1. Badgett, M., Auer, A., Carmichael, L, Parrish, C., and Bull, J. (2002). Evolutionary dynamics of viral attenuation. J. Virol. 76, 10524-9.

2. Beaudry, A., and Joyce, G. (1992). Directed Evolution of an RNA Enzyme. Science 257, 635-41.

3. Cen, S., Khorchid, A., Javanbakht, H., Gabor, J., Stello, T., Shiba, K., Musier-Forsyth, K., and Kleiman, L. (2001). Incorporation of lylsyl-tRNA synthetase into human immunodeficiency virus type 1. J. Virol 75, 5043-8.

4. Cen, S., Javanbakht, H., Kim, S., Shiba, S., Craven, R., Rein, A., Ewalt, K., Schimmel, P., Musier-Forsyth, K., and Kleiman, L. (2002). Retrovirus-specific packaging of aminoacy-tRNA synthetases with cognate primer tRNAs. J Virol 76, 13111-5.

5. Cen, S., and Kleiman, L. (2004). The tRNA-Lys packaging complex in HIV-1. J Biocel 36, 1776-86.

6. Ditzler, M., Bose, D., Shkriabai, N., Marchand, B., Sarafianos, S., Kvaratskhelia, M., and Burke, D. (2011). Broad-spectrum aptamer inhibitors of HIV reverse transcriptase closely mimic natural substrates. Nucleic Acids Res. 1-11.

7. Gabor, J., Cen, S., Javenbakht, H., Niu, M., and Kleiman, L. (2002). Effect of altering the tRNA-Lys,3 concentration in human immunodeficiency virus type 1 upon its annealing to viral RNA, GagPol incorporation, and viral infectivity. J. Virol. 76, 9096-9102.

8. Ganser-Pornillos, B., Yeager, M., and Sundquist, W. (2008). The structural biology of HIV assembly. Curr. Opin. Struct. Biol. 18,203-17.

9. Goldschmidt, V., Paillart, J., Rigourd, M., Ehresmann, B., Aubertin, A., Ehresmann, C., and Marquet, R. (2004). Structural variability of the initiation complex of HIV-1 reverse transcription. J. Biol. Chem. 34, 35923-31.

10. Hall, K., and Kranz, J. (2010). Nitrocellulose filter binding for determination of dissociation constants. Methods in Molecular Biology 118, 105-114.

11. Halwani, R., Cen, S., Javanbakht, H., Saadatmand, J., Kim, S., Shiba, K., and Kleiman L. (2004). Cellular distribution of lysyl-tRNA synthetase and its interaction with Gag during human immunodeficiency virus type 1 assembly. J. Virol. 78, 7553-64.

12. Held, D., Kissel, J., Patterson, J., Nickens, D., and Burke, D. (2006). HIV-1 inactivation by nucleic acid aptamers. Front. Bioscii. 11, 89-112.

13. Javanbakht, H., Halwani, R., Cen, S., Saadatmand, J., Musier-Forsyth, K., Gottlinger, H., and Kleiman, L. (2003). The interaction between HIV-1 Gag and human lysyl-tRNA synthetase during viral assembly. J. Biol. Chem. 30, 27644-51.

14. Jhaveri, S., and Ellington, A. (2000). In vitro selection of RNA aptamers to a protein arget by filter immobilization. Current Protocols in Molecular Biology 24.3.1-24.3.25.

15. Nei, M., Maruyama, T., and Chakraborty, R. (1975). The bottlecneck effect and genetic variability in populations. Evolution 29, 1-10.

16. Ramalingam, D., Duclair, S., Datta, S., Ellington, A., Rein, A., and Prasad, V. (2011). RNA aptamers directed to human immunodeficiency virus type 1 Gag polyprotein bind to the matrix and nucleocapsid domains and inhibit virus production. J. Virol. 85, 305-14.

17. Shi, J., Francklyn, C., Hill, K., and Schimmel, P. (1990). A nucleotide that enhances the charging of RNA minihelix sequence variants with alanine. Biochem. 29, 3621-26.

18. Shiba, K., Stello, T., Motegi, H., Noda, T., Musier-Forsyth, K., and Schimmel, P. (1997). Human lysyl-tRNA synthetase accepts nucleotide 73 variants and rescues *Eschericichia coli* double-defective mutant. J. Biol. Chem. 272, 22809-22816.

19. Stello, T., Hong, M., and Musier-Forsyth, K. Efficient aminoacylation of tRNA-Lys,3 by human lysyltRNA synethetase is dependent on covalent continuity between the acceptor stem and the anticodon domain. Nucl. Acids Res. 27, 4823-29.

20. Tuerk, C., and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249, 505-10.